

VŠB – Technická univerzita Ostrava  
Fakulta elektrotechniky a informatiky  
Katedra kybernetiky a biomedicínského inženýrství

Biometrické softwarové prostředí pro identifikaci, klasifikaci a  
sledování uživatele na základě analýzy obrazu obličeje

Biometric Software Environment for Identification, Classification and  
Tracking of User based on Facial Image Analysis

VŠB - Technická univerzita Ostrava  
Fakulta elektrotechniky a informatiky  
Katedra kybernetiky a biomedicínského inženýrství

## Zadání diplomové práce

Student: **Bc. David Sojka**

Studijní program: N2649 Elektrotechnika

Studijní obor: 3901T009 Biomedicínské inženýrství

Téma: **Biometrické softwarové prostředí pro identifikaci, klasifikaci a sledování uživatele na základě analýzy obrazu obličeje**  
**Biometric Software Environment for Identification, Classification and Tracking of User based on Facial Image Analysis**

Jazyk vypracování: čeština

Zásady pro vypracování:

1. Nastudování problematiky biometrických systémů s důrazem na obrazovou analýzu.
2. Rešerše metod detekce obličeje z obrazu.
3. Rešerše metod sledování a klasifikace obličeje z obrazu.
4. Tvorba testovací báze faciálních obrazů pro testování.
5. Design a realizace algoritmu pro detekci obličeje ze statických a dynamických obrazů.
6. Design a realizace klasifikačních algoritmů pro sledování a rozpoznání obličeje.
7. Realizace softwarového prostředí pro biometrickou identifikaci uživatele na základě zpracování obrazu.
8. Testování SW prostředí na reálných statických a dynamických datech.
9. Objektivní vyhodnocení robustnosti a efektivity navrženého biometrického systému.
10. Zhodnocení výsledků práce.

Seznam doporučené odborné literatury:

- [1] DRAHANSKÝ, Martin, Michal DOLEŽEL a Filip ORSÁG. *Biometrie*. Brno: [M. Drahanský], 2011. ISBN 978-80-254-8979-6.
- [2] MCANDREW, Alasdair. *Introduction to digital image processing with MATLAB*. Boston: Thomson Course Technology, c2004. ISBN 0-534-40011-6.
- [3] NILSSON, Nils J. *Principles of artificial intelligence*. Los Altos: Morgan Kaufmann Publishers, [1980]. ISBN 0-934613-10-9.
- [4] DAOUDI, Mohamed, Anuj SRIVASTAVA a Remco Coenraad VELTKAMP, ed. *3D face modeling, analysis and recognition*. Chichester: Wiley, 2013. ISBN 978-0-470-66641-8.

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí diplomové práce: **doc. Ing. Martin Augustynek, Ph.D.**


Konzultant diplomové práce: **Ing. Jan Kubíček, Ph.D.**

Datum zadání: 01.09.2019

Datum odevzdání: 30.04.2020



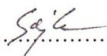
  
\_\_\_\_\_  
doc. Ing. Jiří Koziorek, Ph.D.  
*vedoucí katedry*

  
\_\_\_\_\_  
prof. Ing. Pavel Brandštetter, CSc.  
*děkan fakulty*

## Prohlášení studenta

„Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.“

V Ostravě dne 30. 4. 2020

..........

## **Poděkování**

Rád bych poděkoval vedoucímu mé diplomové práce doc. Martinovi Augustynkovi, Ph.D. za supervizi mé práce. Rovněž bych také chtěl rád touto cestou poděkovat konzultantovi mé diplomové práce Ing. Janu Kubíčkovi, Ph.D. za odborné a cenné rady a směřování aktivit během vypracování této práce.

## **Abstrakt**

Tato diplomová práce se zabývá návrhem a realizací hybridního biometrického systému pro detekci, sledování a klasifikaci obličeje ze statických a dynamických obrazů.

V práci jsou popsány různé metody detekce, sledování a rozpoznání obličejů z obrazových dat. Dále je v práci obsažen postup realizace biometrického systému s využitím algoritmů Viola-Jones, KLT a konvoluční neuronové sítě s předtrénovanou sítí AlexNet. Práce také obsahuje objektivní testování vytvořeného systému vůči variabilním degradačním vlivům a popis tvorby SW prostředí pro účely testování identifikace uživatele.

## **Klíčová slova**

Biometrie, Viola-Jones, AlexNet, Kanade-Lucas-Tomasi, MATLAB, detekce obličeje, sledování obličeje, rozpoznání obličeje, konvoluční neuronová síť

## **Abstract**

This diploma thesis deals with the design and realization of a hybrid biometric system for face detection, tracking and recognition from static and dynamic images.

The thesis describes different methods of face detection, tracking and recognition from 2D data. The thesis involves the process of realizing the biometric system using Viola-Jones, KLT algorithm and pretrained AlexNet convolutional neural network. In addition, the thesis includes also objective testing of created system towards the variable degradation effects and describes the realized SW environment for user classification purposes.

## **Key words**

Biometry, Viola-Jones, AlexNet, Kanade-Lucas-Tomasi, MATLAB, face detection, face tracking, face recognition, convolutional neural network

## Obsah

Seznam použitých zkratk.....	9
Seznam obrázků .....	10
Seznam tabulek .....	13
Úvod.....	14
1 Biometrické vlastnosti a systémy .....	15
1.1 Biometrické vlastnosti a jejich charakteristiky .....	15
1.1.1 Anatomické vlastnosti .....	16
1.1.2 Behaviorální vlastnosti.....	18
1.2 Biometrické systémy .....	20
1.2.1 Unimodální systémy .....	20
1.2.2 Multimodální systémy .....	20
2 Analýza obrazových dat v biometrických systémech .....	22
2.1 Základní obrazové parametry .....	22
2.2 Předzpracování .....	24
2.3 Segmentace obrazu.....	29
3 Metody detekce obličeje z obrazu.....	32
3.1 Metody na základě znalostí o lidské tváři – Knowledge-based methods .....	32
3.2 Metody na základě neměnných rysů obličeje – Feature invariant methods .....	33
3.3 Metody na základě porovnávání se vzorem – Template matching .....	34
3.4 Metody na základě vzhledu – Appearance-based methods.....	36
4 Metody sledování a rozpoznání obličeje z obrazu .....	38
4.1 Face tracking - sledování detekované tváře .....	38
4.1.1 Optical Flow .....	38
4.1.2 Sledování CAMShift.....	39
4.2 Rozpoznání tváře.....	40
4.2.1 Eigenfaces .....	41
4.2.2 Fisherfaces.....	42
4.3 Neuronové sítě .....	43
4.3.1 Hluboké učení – deep learning.....	43
4.3.2 Model neuronu .....	44

4.3.3	Předtrénované konvoluční neuronové sítě .....	45
5	Realizace a implementace algoritmů pro detekci a rozpoznávání obličejů .....	46
5.1	Požadavky na snímanou scénu .....	46
5.2	Komunikace webkamery a MATLABu .....	47
5.3	Tvorba databáze, detekce a sledování obličejů ve videu .....	48
5.3.1	Algoritmus pro tvorbu databáze .....	49
5.3.2	Detekce obličeje metodou Viola-Jones .....	51
5.3.3	Sledování obličeje metodou KLT .....	55
5.4	Realizace systému pro rozpoznání identity člověka na základě obličeje .....	56
5.4.1	Konvoluční neuronová síť – CNN .....	56
5.4.2	AlexNet .....	57
5.4.3	Vrstvy sítě AlexNet .....	57
5.4.4	Učení CNN .....	61
6	Testování a vyhodnocení robustnosti realizovaného biometrického systému .....	64
6.1	Míra přesnosti detekovaných obličejů sledováním a jednotlivou detekcí .....	64
6.2	Objektivní testování robustnosti detekce a klasifikace obličejů .....	65
6.2.1	Spolehlivost detekce a klasifikace vůči změně směru osvětlení scény .....	66
6.2.2	Spolehlivost detekce a klasifikace vůči přítomnosti šumu v obraze .....	67
6.2.3	Spolehlivost detekce a klasifikace vůči změně rozlišení obrazu .....	71
6.2.4	Spolehlivost detekce a klasifikace vůči změně expozice obrazu .....	72
7	Realizace SW prostředí pro biometrickou identifikaci uživatele .....	74
7.1	Panel pro načtení vstupních dat .....	74
7.2	Oblast volitelného nastavení .....	75
7.3	Extrakce obličejů a natrénování sítě .....	76
7.4	Detekce a rozpoznání .....	77
7.5	Degradace testovaného obrazu .....	77
8	Zhodnocení výsledků práce .....	79
	Závěr .....	80
	Zdroje .....	81
	Seznam příloh .....	84
	Příloha I .....	85



Příloha II.....	86
Příloha III .....	87

## **Seznam použitých zkratek**

**2D** – dvourozměrný

**CAMS** – Continuously Adaptive Mean Shift

**CCD** – Charge-coupled device

**CMOS** - Complementary Metal-Oxide Semiconductor

**CNN** – Convolutional Neural Networks

**Dpi** – Dots per inch

**FPS** – Frames Per Second

**GUI** – Graphical User Interface

**HSV** – Hue Saturation Value

**IR** – infrared

**JPEG** – Joint Photographic Experts Group

**KLT** – Kanade-Lucas-Tomasi

**k-NN** – Nearest Neighbours

**LDA** – Linear Discriminant Analysis

**PCA** – Principal Component Analysis

**ReLU** – Rectified Linear Unit

**RGB** – Red Green Blue barevné spektrum

**ROI** – Region of Interest

**SVM** – Support Vector Machines

**SW** - SoftWare

**TP** – Track Point

**USB** – Universal Serial Bus

**YCbCr** – Luminance Chromatic blue Chromatic red

## Seznam obrázků

Obrázek 1: Anatomické vlastnosti člověka .....	16
Obrázek 2: Třídy papilárních linií zleva: oblouk, klenutý oblouk, pravá smyčka, levá smyčka, vír (upraveno) .....	17
Obrázek 3: Snímky na různém rozsahu vlnových délek IR záření, a) viditelné záření, b) krátké vlnové délky IR záření SWIR, c) střední vlnové délky IR záření MWIR, d) dlouhé vlnové délky IR záření LWIR .....	17
Obrázek 4: Oční duhovka (vlevo) a sítnice (vpravo) .....	18
Obrázek 5: Dynamické vlastnosti člověka, zleva hlas, dynamika stisku kláves, podpis .....	19
Obrázek 6: Průběh amplitudy zvukového signálu (slova „Emanuel“) v čase [1] .....	19
Obrázek 7: Multimodální biometrický systém (upraveno podle [1]) .....	21
Obrázek 8: Digitální 2D obraz, zleva: celý obraz, přiblížený obraz, jasové hodnoty pixelů .....	22
Obrázek 9: Reprezentace barevného modelu RGB (upraveno podle [9]) .....	23
Obrázek 10: Transformace pixelů z RGB modelu do monochromatického kanálu (přeupraveno podle [9]) .....	23
Obrázek 11: Barevný model YCbCr .....	24
Obrázek 12: Ekvalizace histogramu. Zleva původní obraz, histogram původního obrazu, obraz po ekvalizaci, ekvalizovaný histogram (vlastní) .....	25
Obrázek 13: Vlevo deformovaný obraz, vpravo korekce obrazu geometrickými transformacemi [10] .....	25
Obrázek 14: Vstupní snímek (vlevo nahoře), vstupní snímek s aditivním šumem „Salt & Pepper“ (vpravo nahoře), snímek po filtraci Gaussovým filtrem 4x4 (vlevo dole), snímek po filtraci mediánovým filtrem 4x4 (vpravo dole), (vlastní) .....	26
Obrázek 15: Typy obrazových hran (vlastní podle [11]) .....	27
Obrázek 16: Detekce hran za pomoci tří různých operátorů. Zleva původní snímek, Robertsův operátor, Cannyho operátor (vlevo dole), Zero-crossing (vpravo dole), (vlastní) .....	28
Obrázek 17: Základní binární morfologické operace – dilatace a eroze [1] .....	29
Obrázek 18: Znázornění segmentace metodou prahování. Zleva původní snímek, binární obraz s prahem $T = 180$ (histogram oranžová značka), multiregionální obraz s prahy $T_1 = 140$ a $T_2 = 220$ (histogram zelená značka) .....	30
Obrázek 19: Histogram s vyznačenými prahy segmentace pro binární a multiregionální segmentaci .	30
Obrázek 20: Možnost vytvoření obrazových bloků reprezentující obličej [14] .....	32
Obrázek 21: Postupné snižování rozlišení zkoumaného snímku [13] .....	33
Obrázek 22: Původní snímek, binární obraz detekce kůže z modelu YCbCr, binární snímek po použití morfologických operací [17] .....	34

Obrázek 23: Referenční vzor obličeje skládající se z 16 subšablon a 23 vzájemných vztahů (11 nezbytných a 12 potvrzujících) [13].....	35
Obrázek 24: Ukázka deformovatelných referenčních vzorů oka a úst.....	36
Obrázek 25: Klasifikátor (vlevo) a klasifikační třídy $\omega_1$ a $\omega_2$ (vpravo) [22].....	36
Obrázek 26: Znázornění jedné iterace algoritmu Mean Shift [32].....	40
Obrázek 27: Obecný postup rozpoznávání obličejů (vlastní) .....	40
Obrázek 28: Znázornění převodu obrazu na řádkový vektor (vlastní).....	41
Obrázek 29: Zpětná transformace obrazu z redukovaného prostoru vlastních obličejů vzniklého metodou PCA (upraveno podle [29]).....	42
Obrázek 30: Vymezení projekce prostoru a vlastních bodů pro dvě třídy metodou PCA a LDA [30].	43
Obrázek 31: Vlastní obličeje vytvořené metodou PCA (nahore) a LDA (dole) [31].....	43
Obrázek 32: Architektura hlubokých neuronových sítí [44].....	44
Obrázek 33: Model neuronu (vlastní) .....	45
Obrázek 34: Postup realizace biometrického systému .....	46
Obrázek 35: Jednotlivé datasety vytvořené databáze obličejů .....	48
Obrázek 36: Rozdělení pořízené databáze na trénovací a testovací databázi .....	49
Obrázek 37: Realizace tvorby databáze (vlastní).....	50
Obrázek 38: Haarova vlnka (vlastní) .....	51
Obrázek 39: Základní typy Haarových příznaků ve čtvercových oknech (vlastní) .....	52
Obrázek 40: Reálné použití Haarových příznaků v obraze [21] .....	52
Obrázek 41: Integrovaný obraz: a) hodnota bodu A reprezentující modře vyznačenou oblast původního obrazu, b) plocha S, reprezentována modrou barvou, vycházející z rovnice 5.4 (vlastní).....	53
Obrázek 42: Kaskádní řazení klasifikátorů (upraveno podle [1]) .....	54
Obrázek 43: Posun extrahovaných bodů metody optical flow (vlastní podle [32]) .....	55
Obrázek 44: Aliasing: vlevo neznámý pohyb, vpravo skutečný pohyb [33] .....	56
Obrázek 45: Architektura konvoluční neuronové sítě [35] .....	57
Obrázek 46: Znázornění konvoluce, jejího jádra a paddingu.....	58
Obrázek 47: Přenosová funkce ReLU [39] .....	59
Obrázek 48: Sdružování funkcí Max Pooling .....	59
Obrázek 49: Proces přerušení spojení neuronů ve vrstvě sítě dropout [40].....	60
Obrázek 50: Klasifikace CNN: FC vrstva, Softmax vrstva, klasifikační třídy (upraveno podle [44]) .	60
Obrázek 51: Gradient descent [42].....	62
Obrázek 52: Proces trénování sítě AlexNet rozšířenou trénovací databází DAT2 .....	63
Obrázek 53: Objektivní testování robustnosti algoritmu vůči různým druhům degradace obrazu .....	65

Obrázek 54: Typy jednotlivých osvětlení zleva: test_dataset_1, test_dataset_2, test_dataset_3, test_dataset_4 .....	66
Obrázek 55: Graf robustnosti biometrického systému vůči osvětlení.....	67
Obrázek 56: Graf robustnosti biometrického systému vůči změně směrodatné odchylky Gaussova šumu .....	68
Obrázek 57: Ukázka degradace vstupního obrazu Gaussovým šumem.....	68
Obrázek 58: Graf robustnosti biometrického systému vůči změně směrodatné odchylky šumu Speckle .....	69
Obrázek 59: Ukázka degradace vstupního obrazu šumem Speckle .....	69
Obrázek 60: Graf robustnosti biometrického systému vůči změně hustoty šumu Salt & Pepper .....	70
Obrázek 61: Ukázka degradace vstupního obrazu šumem Salt & Pepper .....	70
Obrázek 62: Graf robustnosti biometrického systému vůči změně rozlišení testovaného obrazu .....	71
Obrázek 63: Graf robustnosti biometrického systému vůči změně expozice testovaného obrazu.....	72
Obrázek 64: Ukázka změny expozice obrazu offsetem .....	73
Obrázek 65: Design softwarového prostředí ve stavu idle s popisem.....	74
Obrázek 66: Panel pro načtení vstupních dat .....	75
Obrázek 67: Přidání nového subjektu do databáze .....	75
Obrázek 68: Videosekvence pořizující snímky nového subjektu.....	76
Obrázek 69: Volba počtu subjektů, ořez obličejů a natrénování sítě .....	76
Obrázek 70: Nově vytvořené a normalizované snímky v příslušné složce databáze .....	76
Obrázek 71: Oblast programu pro detekci a klasifikaci testovaného obrazu .....	77
Obrázek 72: Testovaný obličej a obličej identifikované osoby.....	77
Obrázek 73: Sekce SW prostředí pro degradaci testovaného snímku.....	78
Obrázek 74: Ukázka SW prostředí v průběhu používání aplikace.....	78

## **Seznam tabulek**

Tabulka 1: Charakteristiky biometrických vlastností.....	15
Tabulka 2: Parametry použité webkamery .....	47
Tabulka 3: Srovnání četnosti detekovaných obličejů sledováním a jednotlivou detekcí obličeje .....	64

## Úvod

Biometrické systémy se v současné době rychle rozvíjejí a velkou zásluhu na tomto faktu má stále se zrychlující a zdokonalující výpočetní technika. Biometrické rozpoznávání používá každý z nás prakticky každý den a jedním z nejvyužívanějších biometrických systémů člověka je právě rozpoznávání osob na základě jejich obličeje. V digitální technice do tohoto oboru spadá počítačové vidění založené na zpracování obrazových dat. Samotné snímání a ukládání kvalitních obrazových informací není v dnešní době žádný problém, horší je to ale s náročností zpracování těchto dat. Proto je potřeba biometrické systémy optimalizovat a přistupovat k nim různými způsoby tak, aby byla jejich funkce co nejefektivnější z pohledu přesnosti a rychlosti.

Cílem této diplomové práce je navrhnout a realizovat hybridní biometrický systém pro rozpoznání obličeje, který bude schopen pracovat s reálnými statickými a dynamickými obrazovými daty. Pro takto realizovaný systém bude potřeba objektivně vyhodnotit jeho kvalitu, a to jak kvalitu detekce obličejů, tak i jeho sledování a výsledného rozpoznání. V neposlední řadě bude potřeba vytvořit jednoduché SW prostředí pro biometrickou klasifikaci uživatele.

V teoretické části této práce je zaměřena pozornost zejména na biometrické vlastnosti, jejich charakteristiky a typy biometrických systémů. Dále se pak práce ubírá směrem zpracování digitálních obrazových dat a jejich základními parametry. V kapitolách 3 a 4 jsou rozebrány různé metody a přístupy pro detekci a rozpoznání lidského obličeje včetně popisu klasifikace dat a strojového učení.

Praktická část se zabývá návrhem a realizací biometrického systému a popisem implementovaných metod, mezi které se řadí zejména kaskádní klasifikátor Viola-Jones pro detekci obličeje, algoritmus KLT pro sledování obličejových příznaků ve videu a konvoluční neuronová síť sloužící pro klasifikaci jedinců. Praktická část práce také zahrnuje postup tvorby vlastní databáze obličejů a objektivní otestování systému vůči různým degradacím obrazu. Testované obrazy byly znehodnocovány třemi aditivními šumy, změnou směru osvětlení snímané scény, změnou rozlišení a změnou expozice. Poslední část práce se zabývá tvorbou a popisem SW prostředí pro biometrickou klasifikaci uživatele na základě obrazu obličeje a vyhodnocením efektivity systému.

# 1 Biometrické vlastnosti a systémy

Tato kapitola se zabývá problematikou biometrických systémů, přičemž hlavní pozornost je věnována analýze obrazu. Dennodenně každý z nás používáme biometrické rozpoznávání na základě různých biometrických vlastností. Vzhledem k tomu, že je každý člověk jedinečný, a to jak jednotlivými částmi těla, tak i způsobem chování, můžeme tyto informace zaznamenávat a poté strojově zpracovávat. Úkolem biometrie tedy je automatizovaně rozpoznávat lidské jedince na základě jejich charakteristických anatomických rysů jako je např. obličej, otisk prstu, duhovka oka atd.

## 1.1 Biometrické vlastnosti a jejich charakteristiky

Biometrická vlastnost je rys, vlastnost nebo akce člověka, která se vyznačuje svou unikátností a tudíž může sloužit k nepřebornému množství možností. Mezi nejčastější využití těchto vlastností se řadí zabezpečovací systémy s automatizovaným přístupem nebo k rozpoznávání jedinců v kriminalistice. U systémů pracujících s automatizovaným přístupem dokáží tyto vlastnosti nahradit hesla nebo předměty k identifikaci člověka. Nejběžnějším zabezpečovacím prvkem jsou stále různé druhy hesel a autentizační předměty, avšak vývoj biometrických systémů je stále rozšiřujícím se odvětvím, které mnohonásobně šetří čas a přispívá ke komfortu uživatele. Biometrická autentizace je rychlá a velice spolehlivá metoda, která je už navíc relativně finančně dostupná. Její hlavní výhodou je, že charakteristické biometrické znaky jsou z časového hlediska téměř neměnné, nelze je odcizit nebo zapomenout [5].

Mezi hlavní kritéria při pořizování konkrétních biometrických systémů je potřeba brát ohled na charakteristiky biometrických vlastností. Existuje osm základních charakteristik se třemi stupni hodnocení (V – vysoká, S – střední, N - nízká). Všechny základní charakteristiky jsou uvedeny v tabulce pro vybrané biometrické vlastnosti (Tabulka 1). Krom těchto charakteristik je nutno také zohlednit cílovou skupinu, náklady spojené se systémem a míru požadované bezpečnosti [1] [7].

Tabulka 1: Charakteristiky biometrických vlastností

Charakteristika Biom. vlastnost	Univerzalita	Unikátnost	Neměnnost	Dostupnost	Výkonnost	Akceptace	Bezpečnost	Finanční náklady
Obličej	V	N	S	V	N	V	N	N
Otisk prstu	S	V	V	S	V	S	V	N
Geometrie ruky	S	S	S	V	S	S	S	S
Duhovka	V	V	V	S	V	N	V	V
Sítnice	V	V	S	N	V	N	V	V
Podpis	N	N	N	V	N	V	N	N
Hlas	S	N	N	S	N	V	N	N
Termogram	V	V	N	V	S	V	V	V
DNA	V	V	V	N	V	N	N	V

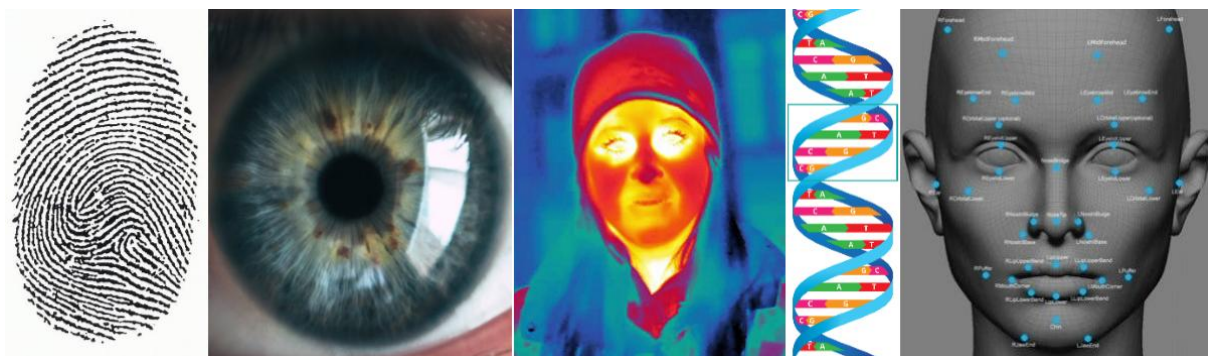


Biometrické vlastnosti se podle definice rozdělují do dvou kategorií, a to do anatomických a behaviorálních.

### 1.1.1 Anatomické vlastnosti

U těchto vlastností je jeden pevný rys člověka vždy jednou biometrickou vlastností, která není lehce ovlivnitelná stavem člověka. Metoda analýzy anatomických vlastností se také nazývá jako statická metoda [3]. Mezi anatomické vlastnosti patří:

- obličej
- termogram obličeje
- otisk prstu
- sítnice oka
- duhovka oka
- geometrie ruky
- dlaň
- dentální obraz
- tvar ucha
- DNA



Obrázek 1: Anatomické vlastnosti člověka

### Otisky prstů

Identifikace na základě otisků prstů je v dnešní době asi nejznámější a nejvyužívanější metodou klasifikace osob. Každý prst má totiž jedinečně tvarované papilární linie (vyvýšené reliéfy kůže), které lze jednoduše rozlišit a klasifikovat. Papilární linie jsou také relativně neměnné v čase, a proto je rozpoznávání jedinců velmi spolehlivé. Jediným poškozením této informace jsou vnější vlivy, jako je např. pořezání nebo různé poruchy kůže. Obor studující otisky prstů je nazýván daktyloskopií a jeho hlavní využití je v kriminalistice.

Papilární linie prstů mají výšku v rozmezí od 0,1 do 0,4 mm a šířku od 0,2 do 0,5 mm. Tyto linie tvoří na břišku prstu vzory, které jsou klasifikovány do několika základních tříd. Mezi nejčastější vzory patří oblouk, klenutý oblouk, vír, levá smyčka a pravá smyčka (Obrázek 2) [1].

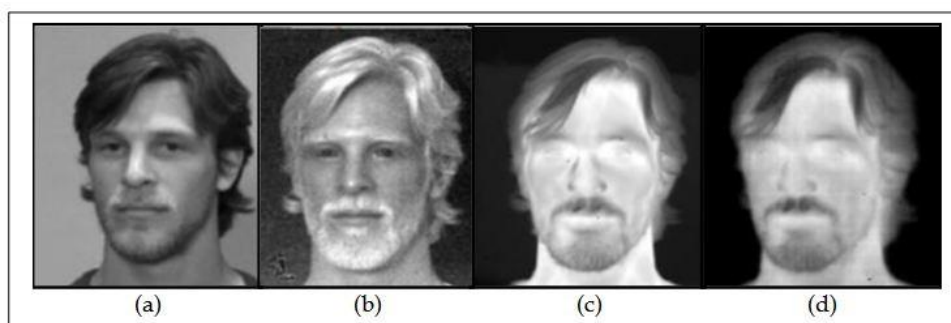


Obrázek 2: Třídy papilárních linií zleva: oblouk, klenutý oblouk, pravá smyčka, levá smyčka, vír (upraveno)

Pro získání této digitální informace se využívají snímače otisků prstů s rozlišením od 250 do 1000 dpi (dots per inch). Další parametrem je počet bitů pro zakódování barvy, standardně 8 bitů pro odstíny šedé barvy. Otisky prstů se nejčastěji snímají pomocí optických CCD (Charge-Coupled Device) nebo CMOS (Complementary Metal-Oxide Semiconductor) snímači, nebo také lineárními kapacitními senzory (změnou kapacity v místě dotyku). Takto získaný digitální snímek je pak potřeba za pomoci různých algoritmů vyhodnotit a jednoznačně klasifikovat osobu, jež je jeho autorem. V dnešní době je tato metoda velice rychlá a její spolehlivost přesahuje 99,5 % [2].

### Termogram obličeje

Obraz obličeje nemusí být snímán pouze ve viditelném spektru elektromagnetického záření. Lze jej snímat také na vlnových délkách odpovídajících infračervenému záření. Takovým snímkům se říká termogramy a jsou pořizovány termokamerou. Tyto snímky nesou informace o vyzařovaném množství záření tělesem v závislosti na jeho teplotě (resp. pohybu částic v krystalické mřížce). Je také nutné podotknout, že na různém rozsahu vlnových délek elektromagnetického záření, získáváme různé obrazy, viz Obrázek 3. U snímání termogramů se využívá různých parametrů, mezi které patří např. emisivita (pro lidskou kůži  $\epsilon = 0,98$ ), teplota okolí, vzdálenost od objektu a vlhkost okolí. Největší výhodou této metody oproti klasickým snímkům obličeje je nezávislost na osvětlení a také velmi složitá možnost falsifikace. Na druhou stranu, termogram obličeje značně závisí na teplotě okolí, fyzické aktivitě člověka nebo také emocích. Proto je vhodné využívat tuto metodu v kombinaci s dalšími metodami (klasické 2D snímky), kdy dochází k mnohem přesnějším výsledkům.

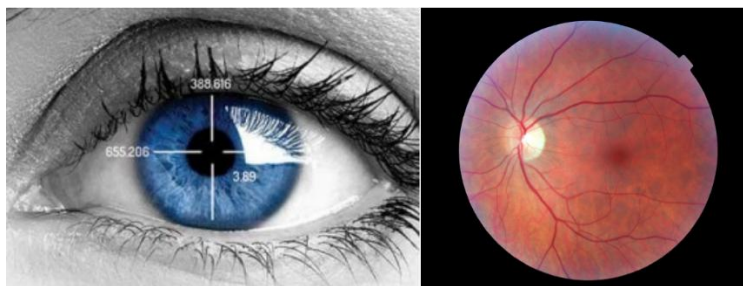


Obrázek 3: Snímky na různém rozsahu vlnových délek IR záření, a) viditelné záření, b) krátké vlnové délky IR záření SWIR, c) střední vlnové délky IR záření MWIR, d) dlouhé vlnové délky IR záření LWIR

## Vlastnosti oka

V lidském oku je také možné sledovat unikátní biometrické vlastnosti, jako jsou duhovka a sítnice. Výhodou těchto biometrických vlastností je především větší zabezpečení proti falsifikaci, z důvodu jejich složitějšího pořízení a také k mnohonásobně větší biometrické entropii (vzorkování duhovky je unikátní i pro jednovaječná dvojčata). Snímek oční duhovky se pořizuje nejčastěji CCD snímačem, je monochromatický a jeho pořízení je neinvazivní. Doba identifikace oční duhovky se pohybuje okolo 2 sekund. Další výhodou sledování oční duhovky je to, že v průběhu života nemění svou základní strukturu. Na druhou stranu, pro oklamání detektoru bez detekce živosti stačí získat fotografii duhovky. Nejčastějším detektorem živosti u duhovky je reakce na změnu osvětlení [1], [5].

Při získávání snímku oční sítnice je na rozdíl od předcházející metody zapotřebí složitějšího vybavení, a to proto, že se zde využívá IR (infračerveného) laserového paprsku směřujícího přímo do oka a snímáním jeho odrazu od sítnice (retinoskopy a fundus kamery). Z praktického hlediska je tato metoda více omezená než u duhovky, kdy je zapotřebí získat dostatečně kvalitní snímek očního pozadí, avšak její obelstění je téměř nemožné a časová náročnost se pohybuje okolo 2 až 4 sekund [1], [5].

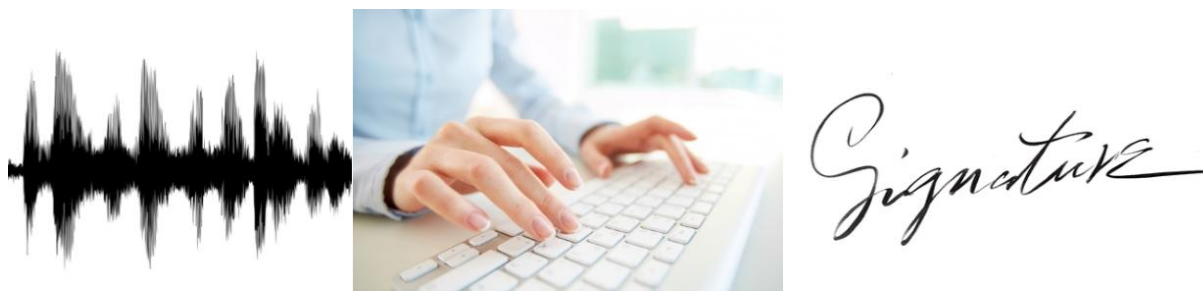


*Obrázek 4: Oční duhovka (vlevo) a sítnice (vpravo)*

### 1.1.2 Behaviorální vlastnosti

Behaviorální vlastnosti jsou často nazývány taky jako dynamické vlastnosti a to proto, že jsou spojeny s nějakou akcí člověka, která je jistým způsobem jedinečná. Tyto biometrické vlastnosti jsou velmi lehce ovlivnitelné, a proto se často stává, že se nasnímané vzorky liší. Jedná se tedy o dynamickou metodu analýzy a její náročnost je značně složitější než u metody statické. Mezi základní dynamické vlastnosti patří:

- hlas
- chůze
- mimika obličeje
- dynamika stisku kláves
- písmo a podpis



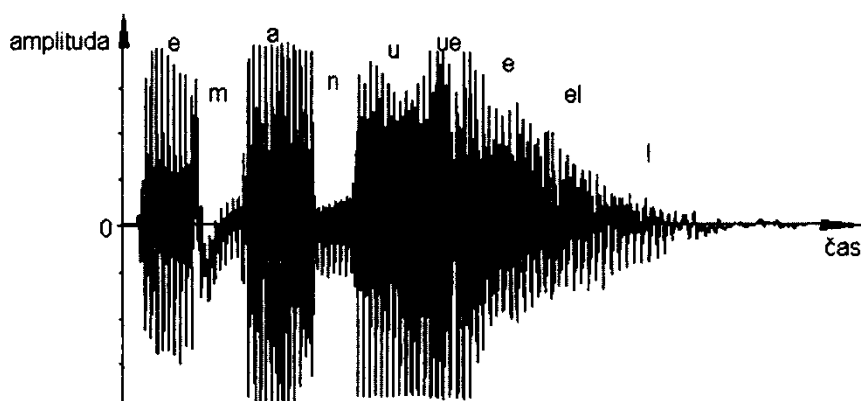
Obrázek 5: Dynamické vlastnosti člověka, zleva hlas, dynamika stisku kláves, podpis

## Rozpoznání hlasu

Rozpoznávání lidského hlasu a strojové porozumění lidské řeči je velmi složitý úkol, který stále vyžaduje velké množství práce a studií. Rozpoznání řeči je schopnost stroje nebo programu přijímat mluvené slovo (Obrázek 6), porozumět mu a na jeho základě vykonávat příkazy. Rozpoznávání hlasu získalo význam a reálné použití se vzestupem umělé inteligence a strojového učení na bázi neuronových sítí, jako například Amazon Alexa, Apple Siri a Microsoft Cortana.

Software pro rozpoznávání hlasu v počítačích vyžaduje převod analogového signálu na digitální signál. Aby byl počítač schopný hlas dešifrovat, musí mít jednak databázi slov nebo slabik, ale také i rychlý prostředek pro porovnávání těchto dat se vstupními signály. Vzory řeči jsou při spuštění programu načítány do paměti RAM. Program rozpoznávání hlasu běží mnohokrát rychleji, pokud lze do RAM načíst celou slovní zásobu, ve srovnání s hledáním některých slov na pevném disku.

Zatímco se míra přesnosti rozpoznávání zlepšuje, všechny systémy a programy rozpoznávání hlasu dělají chyby a hlavním důvodem je to, že lidská řeč je příliš různorodá. Nejtěžším úkolem tedy je nalezení takových parametrů hlasového signálu, které by byly schopny zaručit spolehlivé rozpoznání hlasu dané osoby a jeho význam. Proto jsou tyto metody velmi často založeny na neurčitosti a pravděpodobnosti, tzn. že obsah řeči je nejprve určen přibližně a následně je jeho význam zpřesňován na základě statistických parametrů. Dále také hluk v pozadí může způsobit falešný vstup nebo je zde také problém se slovy, která zní stejně, ale píšou se odlišně. Tento problém může být někdy překonán pomocí uložených kontextových informací. To však bude vyžadovat lepší optimalizace programu, více paměti RAM a rychlejší procesory, než jaké jsou v současné době k dispozici v osobních počítačích [4].



Obrázek 6: Průběh amplitudy zvukového signálu (slova „Emanuel“) v čase [1]

Lidský hlas je jedinečný, avšak velmi lehce ovlivnitelný, proto se u zabezpečovacích systémů používá především v kombinaci s jinými systémy vyhodnocujícími více biometrických vlastností současně (multimodální systémy, viz kapitola 1.2).

## **1.2 Biometrické systémy**

Jak již bylo naznačeno v předchozí kapitole, biometrické systémy mají za úkol rozpoznávat anatomické a behaviorální vlastnosti člověka a na jejich základě zpracovávat příkazy. V této práci bude dále ubírána pozornost především na rozpoznávání lidského obličeje, tedy na anatomickou vlastnost člověka z digitálního 2D obrazového signálu, viz kapitola 2.

Jedním z nejdůležitějších faktorů u biometrických systémů je spolehlivost. Ta se konkrétně u rozpoznávání obličeje může měnit v závislosti na mnoha vnějších vlivech, jako jsou např. osvětlení, zranění, make-up, nasazení brýlí, nárůst vousů, úprava vlasů, pokrývka hlavy nebo stárnutí. Proto je vhodné dbát na tyto faktory zřetel a zohlednit také prostředí a cílovou skupinu uživatelů tohoto systému. To platí pro všechny biometrické systémy, v závislosti na cílové skupině a biometrické vlastnosti, se kterou se pracuje. Biometrické systémy se rozdělují do dvou kategorií podle počtu zpracovávaných biometrických vlastností, tj. unimodální a multimodální systémy.

### **1.2.1 Unimodální systémy**

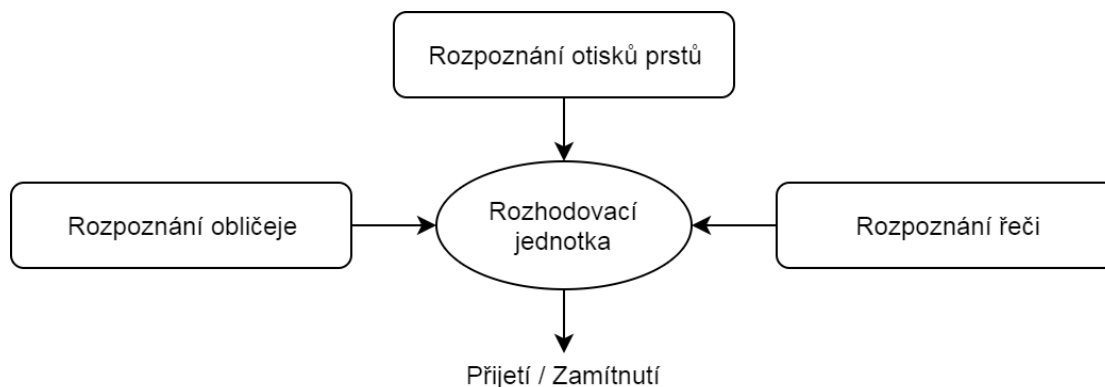
Unimodální systémy pracují pouze s jednou biometrickou vlastností a jejich použití v praxi je stále velmi časté. Slouží k identifikaci nebo verifikaci osoby na základě pouze jedné biometricky s referenčním údajem v databázi. Zpracovává se buď jedna anatomická vlastnost, nebo jedna dynamická vlastnost. Nižší pořizovací náklady těchto systémů úzce souvisí s nižší mírou spolehlivosti. Jednoduchost těchto systémů má za následek častější prolomení bezpečnostního klíče nežádoucí osobou. Mezi další problémy související s těmito systémy se řadí například poranění prstů, změna hlasu při nachlazení, pokrývky hlavy při detekci obličeje a další. Řešením těchto problémů a zajištění větší spolehlivosti identifikace je využití multimodálních systémů.

Příkladem unimodálního systému může být např. samostatné vyhodnocování otisku prstu, dlaně, obličeje, duhovky, sítnice atd.

### **1.2.2 Multimodální systémy**

Multimodální systémy využívají buď více anatomických vlastností (např. rozpoznání obličeje spolu s rozpoznáním duhovky) nebo kombinaci anatomických a dynamických biometrických vlastností. Oproti unimodálním systémům jsou tyto systémy daleko spolehlivější, ale také výpočetně náročnější a finančně nákladnější.

Obrázek 7 znázorňuje příklad multimodálního systému skládajícího se ze tří různých subsystémů, pro jednotlivé biometrické vlastnosti. Systém vyhodnocuje kombinaci tří vstupních anatomických a dynamických vlastností, jako je v tomto případě obličej spolu s otisky prstů a řečí. Na základě všech zjištěných parametrů ze subsystémů, putují informace do tzv. rozhodovací jednotky, ve které systém pomocí vhodného algoritmu rozhodne, zda se jedná o daného jedince či nikoli.



*Obrázek 7: Multimodální biometrický systém (upraveno podle [1])*

Způsobů jak získat větší množství vstupních dat existuje mnoho. Podle zdroje [6] se způsoby dělí na 4 základní typy:

- a) Jedna biometrická vlastnost snímána vícero snímači (např. kombinace 2D a 3D snímku obličeje)
- b) Jedna nasnímaná biometrická informace zpracovaná více metodami rozpoznávání
- c) Větší počet jednoho druhu biometrického znaku (např. otisky všech prstů na ruce)
- d) Kombinace více vlastností dané osoby (Obrázek 7)

Každý z těchto způsobů vyžaduje jiné zpracování vstupních signálů a různé metody rozhodovacího procesu. Pro rozpoznávání obličeje se využívá obrazové analýzy digitálního signálu, které je detailně rozebráno v nadcházející kapitole.

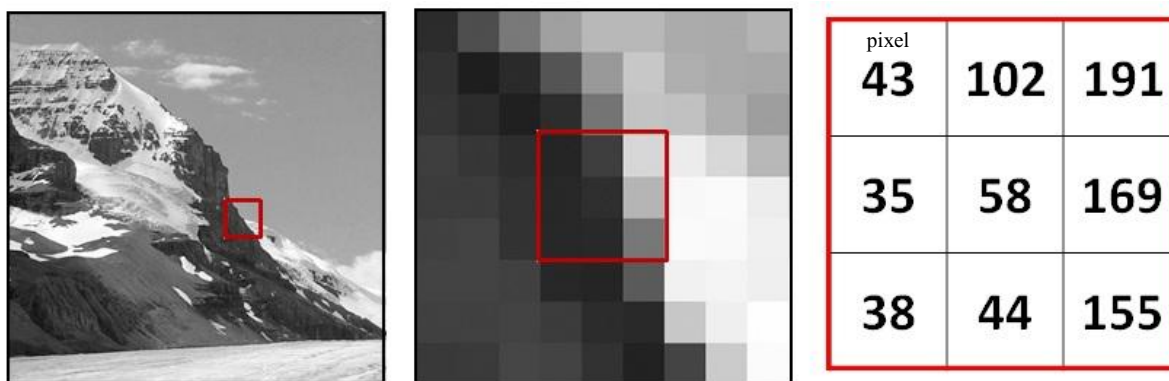


## 2 Analýza obrazových dat v biometrických systémech

Hlavním cílem této práce je rozpoznání obličeje a jeho identifikace z obrazových dat. Celá tato problematika se zabývá zpracováním digitálního 2D signálu. Aby bylo možné osoby podle obličeje identifikovat, je potřeba vědět, z čeho se obraz skládá a jak si jej předpřipravit k následnému zpracování danými metodami.

### 2.1 Základní obrazové parametry

Digitální 2D obraz se skládá z pixelů uspořádaných do dvourozměrné matice (Obrázek 8) o daném rozměru neboli rozlišení. Rozlišením se tedy rozumí počet pixelů ve vertikálním a horizontálním směru. Pixel je nejmenší jednotkou digitálního obrazu, který si nese svou hodnotu intenzity v daném rozsahu určeném bitovou hloubkou, tzn. že každý pixel má konečný počet jasových intenzit. Speciálním případem monochromatického obrazu je obraz binární, kdy pixel dokáže reprezentovat pouze dvě jasové úrovně, a to černou a bílou.



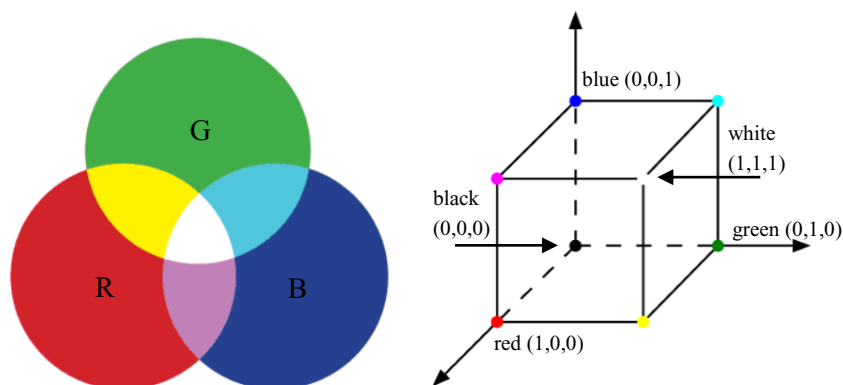
Obrázek 8: Digitální 2D obraz, zleva: celý obraz, přiblížený obraz, jasové hodnoty pixelů

#### Kontrast a dynamický rozsah

Kontrast obrazu je veličina spojená s fyziologií vidění a jde o veličinu, která charakterizuje jeden konkrétní obraz. Kontrast reprezentuje podíl jasových úrovní pixelů v obraze (přesněji hodnotu mezi nejsvětějšími a nejtmavšími místy v obraze) a řadí se k základním obrazovým operacím. Podíl intenzity nejsvětějšího a nejtmavšího pixelu v obraze, který je možné v obraze rozlišit je nazýván jako dynamický rozsah, který je udáván v hodnotách po převodu dekadickým logaritmem [8].

#### Barevný model RGB

V případě barevného 2D obrazu se jedná o složení více chromatických složek z barevného spektra, kde pixel představuje intenzitu barvy jednotlivých vrstev RGB (červená, zelená, modrá). Počet odstínů barev v obraze je tedy dán v závislosti na bitové hloubce jednotlivých RGB spekter a jejich vzájemnou kombinací. Výsledná barva je tedy dána aditivním mícháním všech tří barev, proto se tomuto modelu také říká aditivní (Obrázek 9). S rostoucí intenzitou jednotlivých barevných složek se stává výsledný pixel světlejším (Obrázek 9 vpravo – white (1,1,1)). Hodnoty uvedené v závorce nabývají hodnot 0 až 1 a představují míru intenzity daného spektra, kde 0 reprezentuje nezastoupení dané složky v pixelu (nezastoupením ani jedné ze tří složek vzniká barva černá). Tyto hodnoty se následně přepočítávají konečným počtem možných intenzit v závislosti na bitové hloubce barevných spekter. Standardně se používají 8 bitové vrstvy, což představuje 256 možných intenzit každého spektra. Vzhledem k tomu, že v modelu RGB se míchají 3 vrstvy, vzniká nám přibližně 16 milionů barevných kombinací.

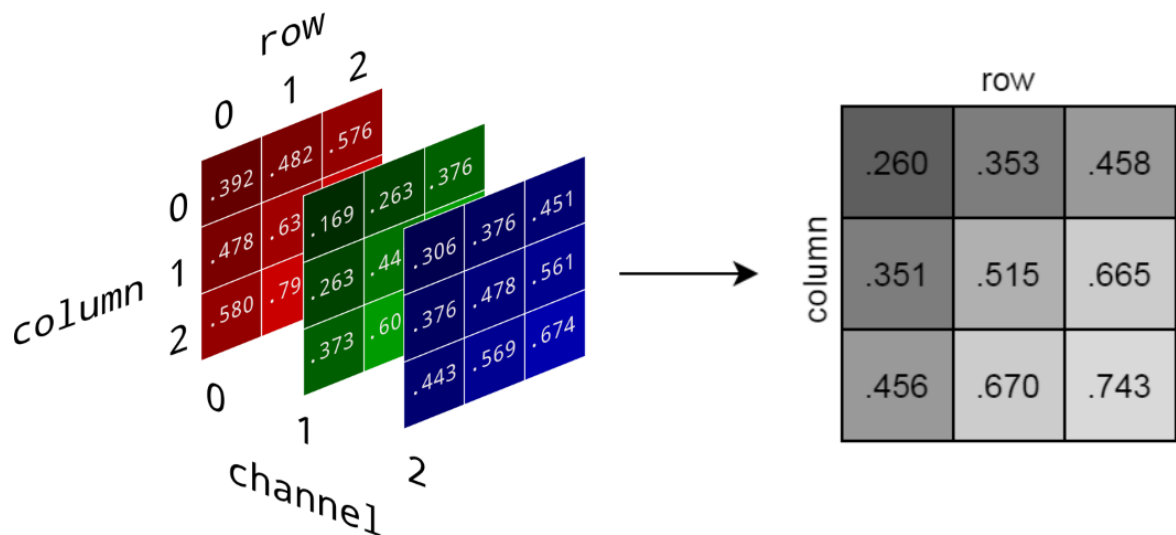


Obrázek 9: Reprezentace barevného modelu RGB (upraveno podle [9])

Pro ušetření množství dat, optimalizaci a zkrácení výpočetního výkonu se často v obrazové analýze pracuje s monochromatickými obrazy. Pro převod obrazu z RGB modelu na šedotónový snímek se využívá následujícího vztahu (2.1), a to z důvodu, že lidské oko vnímá jednotlivé barevné složky s různou intenzitou a tudíž prostým zprůměrováním všech tří kanálů nedostaneme tak kvalitní výsledný monochromatický snímek. Vzhledem k tomu, že lidské oko je nejvíce citlivé na barvu zelenou, tak právě tato barva má na přepočtu největší podíl [9].

$$I = 0,299R + 0,587G + 0,114B \quad (2.1)$$

Znázornění konverze barevného obrazu na obraz monochromatický s využitím předcházejícího vztahu je znázorněn na Obrázek 10.



Obrázek 10: Transformace pixelů z RGB modelu do monochromatického kanálu (přepraveno podle [9])

Veškeré další úpravy v obraze následně provádíme za pomoci maticových a případně vektorových operací. Obrazová analýza se skládá z několika kroků, které jsou nezbytné ke správné identifikaci uživatele. Mezi základní části se řadí předzpracování, segmentace, popis vzorových objektů, detekce objektů a jejich klasifikace.



## Barevný model YCbCr

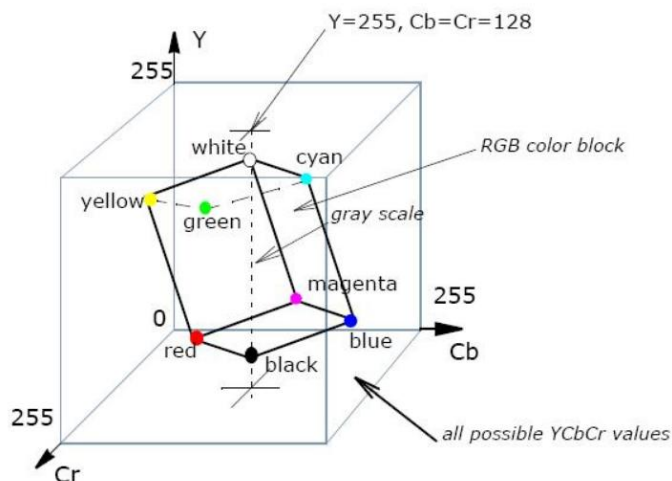
YCbCr (zkratka pro Luminance Chromatic blue Chromatic red) je kódovaný nelineární RGB model, který je běžně používán evropskými televizními standardy anebo slouží také pro kompresi obrazu (kompresní algoritmy JPEG). YCbCr model (znázorněný na Obrázek 11) se skládá ze tří složek, kde Y zastupuje jasovou složku, Cb modrou chrominační složku a Cr červenou chrominační složku. Barevné složky Cb a Cr se využívají v rozsahu -0,5 až 0,5. Jednotlivé složky se přepočítávají z RGB modelu za pomoci následujících rovnic [16] [17]:

$$Y = 0 + 0,299 \cdot R + 0,587 \cdot G + 0,114 \cdot B \quad (2.2)$$

$$Cb = 128 - 0,169 \cdot R - 0,331 \cdot G + 0,500 \cdot B \quad (2.3)$$

$$Cr = 128 + 0,500 \cdot R - 0,419 \cdot G - 0,081 \cdot B \quad (2.4)$$

Lze si povšimnout, že rovnice (2.2) je shodná s rovnicí (2.1) při převodu RGB na odstíny šedi, jelikož se u obou případů jedná o monochromatický snímek. Jediným rozdílem je, že u barevného modelu YCbCr tento snímek představuje pouze jednu složku tohoto modelu.

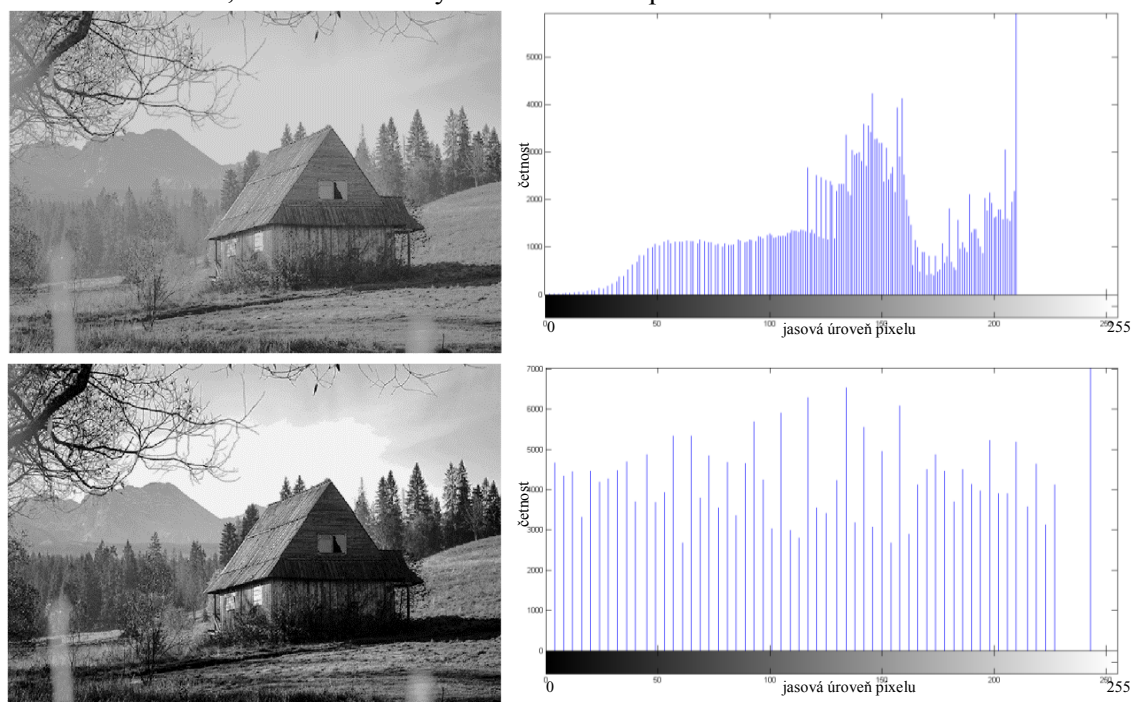


Obrázek 11: Barevný model YCbCr

## 2.2 Předzpracování

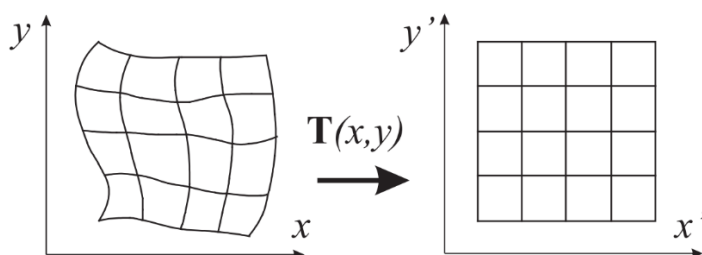
Předzpracování obrazu je nedílnou součástí obrazové analýzy a to z důvodu, že vstupní obraz může být zkreslený, zašuměný nebo jinak deformovaný. Proto je potřeba tyto nežádoucí vlivy z obrazu co nejvíce potlačit. Šum vzniká v obraze vždy a je dán mnoha faktory. Zejména kvalitou snímáče, světelností objektivu a ohniskovou vzdáleností objektivu, která může způsobit zakřivení obrazu. Mezi hlavní operace, zvýrazňující důležité informace obrazu, patří například:

- **Jasové transformace** – úprava histogramu obrazu, zejména jeho ekvalizace. Histogram představuje četnost výskytu jednotlivých hodnot jasu v obraze. Hodnota jasu jednotlivých pixelů nabývá hodnot 0 (nejtmavší - černá) až 255 (nejsvětlejší - bílá). Vzhledem k tomu, že jasové hodnoty pixelů vlivem světelných podmínek nemusí být rozprostřeny vždy po celém rozsahu jasové stupnice (Obrázek 12 – původní obraz a jeho histogram napravo), je potřeba obraz ekvalizovat. To znamená, že veškeré hodnoty jasu pixelů jsou přepočítány tak, aby pokrývaly co největší škálu hodnot jasu tj. od 0 do 255 (Obrázek 12 – ekvalizovaný obraz a jeho histogram vpravo dole). Touto metodou dosáhneme lepšího kontrastu ve snímku a zvýrazníme důležité informace, které se budou využívat v dalším zpracování.



Obrázek 12: Ekvalizace histogramu. Zleva původní obraz, histogram původního obrazu, obraz po ekvalizaci, ekvalizovaný histogram (vlastní)

- **Geometrické transformace** – úpravy geometrických vlastností obrazu. Jak už bylo zmíněno výše, ohniskovou vzdáleností objektivu může vznikat zkreslení obrazu, které můžeme za pomoci geometrických transformací opravit (Obrázek 13). Mezi tyto operace patří také otočení, posunutí a další. V biometrických systémech se geometrické transformace využívají především k normalizaci obrazů, kdy se například nalezené obličeje v obraze zarovnávají na střed nového obrazu pro porovnávání se snímky uloženými v databázi a snazší posouzení jejich podobnosti [10].

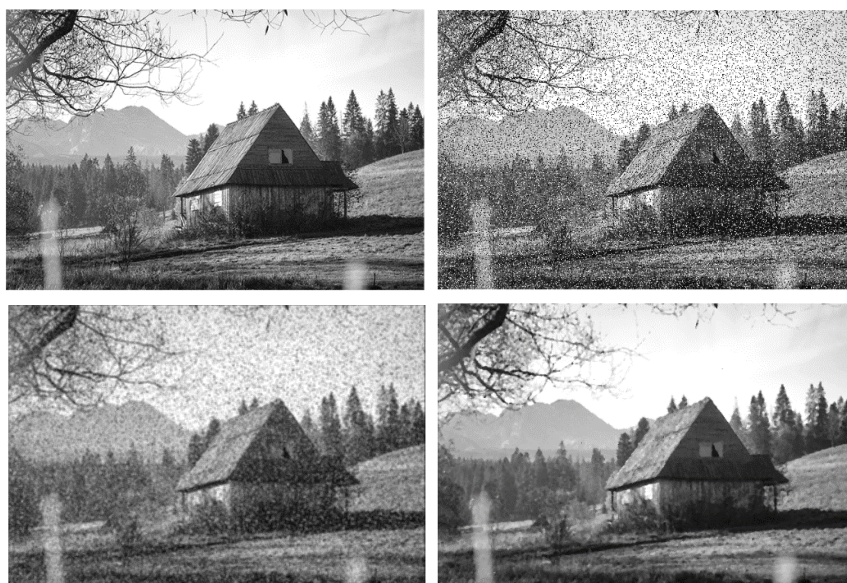


Obrázek 13: Vlevo deformovaný obraz, vpravo korekce obrazu geometrickými transformacemi [10]

- **Filtrace** – redukce šumu vzniklého samotným pořízením snímku, jeho přenosem nebo kompresí.

**Šum**, tento pojem v obrazových informacích, označuje přítomnost nežádoucích dat, která nemají pro pozorovatele žádnou informační hodnotu. Jedná se o nechtěná data zkreslující užitečný signál. Šum může vznikat různými způsoby a podle jeho vlastností jej rozdělujeme do tří základních typů: aditivní šum (Obrázek 14), multiplikativní a kvantizační šum [1]. Za aditivní šum považujeme taková data v obraze, která nejsou žádným způsobem závislá na snímaném signálu (různé zdroje rušení během snímání dat), kdežto šum multiplikativní je ovlivňován právě našim snímaným užitečným signálem. Oba tyto typy šumu lze potlačit filtrací. Speciálním případem je šum kvantizační, který vzniká v důsledku nedostatku kvantizačních hladin během digitalizace signálu nebo špatným zpracováním obrazu. Tento jev lze sledovat v oblasti konstantní barvy na obloze na obraze po ekvalizaci (Obrázek 12). Takto vzniklý šum již nelze zpětně rekonstruovat.

**Filtrace** nejčastěji pracuje v prostorové oblasti obrazu, kde je na vstupní obraz přiložena maska filtru, která se po obraze posouvá. Na základě dané matematické operace, konvoluce, přepočítává hodnoty jednotlivých pixelů a ukládá nové hodnoty pixelů do výstupního obrazu. Jedněmi z nejčastěji používaných filtrů pro potlačení šumu v obraze jsou Gaussův průměrový filtr nebo mediánový filtr. Ukázka těchto základních filtrů je znázorněna obrázkem níže (Obrázek 14).



*Obrázek 14: Vstupní snímek (vlevo nahoře), vstupní snímek s aditivním šumem „Salt & Pepper“ (vpravo nahoře), snímek po filtraci Gaussovým filtrem 4x4 (vlevo dole), snímek po filtraci mediánovým filtrem 4x4 (vpravo dole), (vlastní)*

- **Detekce hran** – zvýraznění místa v obraze s výraznou změnou jasu nebo barvy.

**Hrany** jsou nejzákladnějšími rysy obrazu, podle kterých jsme schopni rozpoznávat jednotlivé objekty. Obrazová hrana je dána vlastnostmi obrazového elementu a jeho okolí, a vyjadřuje směr největšího růstu  $\psi$  a rychlost změny obrazové funkce  $f(x,y)$  – gradient  $\nabla$  představující vektor parciálních derivací podle nezávislých proměnných délky obrazové funkce (rovnice 2.5).

$$\nabla f(x,y) = \left( \frac{\delta f(x,y)}{\delta x}, \frac{\delta f(x,y)}{\delta y} \right) \quad (2.5)$$

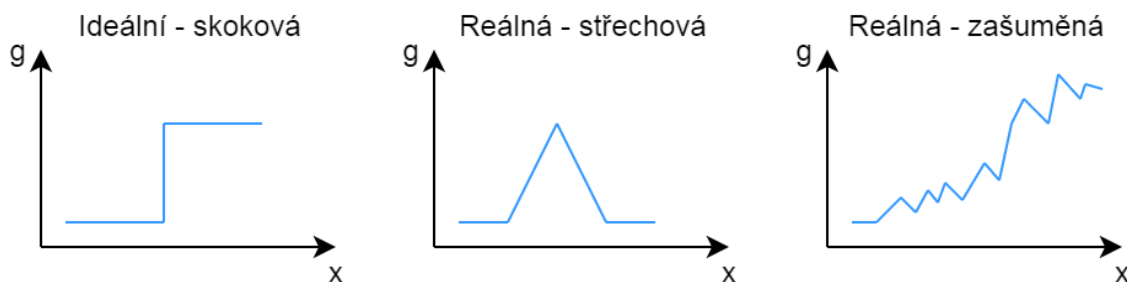
U funkce mající dva parametry, se určuje její gradient velikostí (rovnice 2.6). Tato velikost není závislá na natočení souřadnicového systému ani na směru gradientu, který je dán jediným úhlem  $\psi$  (rovnice 2.7) [1] [11].

$$\|\nabla f(x,y)\| = \sqrt{\left(\frac{\delta f(x,y)}{\delta x}\right)^2 + \left(\frac{\delta f(x,y)}{\delta y}\right)^2} \quad (2.6)$$

$$\psi = \arctg\left(\frac{\delta f(x,y)}{\delta y} / \frac{\delta f(x,y)}{\delta x}\right) \quad (2.7)$$

Jasový profil ideální hrany má skokový průběh. Reálnou hranu v obraze může reprezentovat zidealizovaný střechový průběh nebo také průběh, který je značně ovlivněný přítomností šumu v okolí hranového bodu, viz Obrázek 15.

**Hranový bod** je místo v obraze, které je reprezentováno strmostí gradientu  $\nabla$  funkce  $f(x,y)$  [11].



Obrázek 15: Typy obrazových hran (vlastní podle [11])

Hrany můžeme zvýraznit i pouhým zvýšením kontrastu, avšak pro detailní a přesné detekování hran je nutné využít specifických algoritmů, sloužících tomuto účelu. Samotná detekce a rozpoznání obličejů se řídí hranami, udávající rysy tváře. Hrany mohou být také nesprávně detekovány a to například nesprávným nasvícením objektu, malým osvětlením, nebo šumem přítomným v obraze. Využití hranové detekce je nezbytnou součástí pro následnou segmentaci obrazu. Dalším využitím této detekce je, když potřebujeme obraz zjednodušit, nebo spíše redukovat množství jeho dat. Ponecháním pouze nejpodstatnějších míst v obraze, tedy hran a odstraněním přebytečných dat docílíme toho, že charakter objektu v obraze se nezmění a přesto ušetříme výpočetní výkon a množství dat pro další zpracování. Na detekci hran v obraze často navazují morfologické operace.

Rozlišují se tři základní metody detekce hran v obraze [11].

#### a) Hledání maximálních hodnot prvních derivací

Do této skupiny patří takové gradientní operátory, které využívají maximálních hodnot gradientu po první derivaci obrazové funkce. Vypočtený gradient je následně porovnáván s předem stanoveným prahem a tímto základě probíhá rozhodování, zda dané místo je obrazovou hranou či nikoli. Takovéto operátory využívají konvolučních jádra s různou velikostí a směrem. Mezi základní



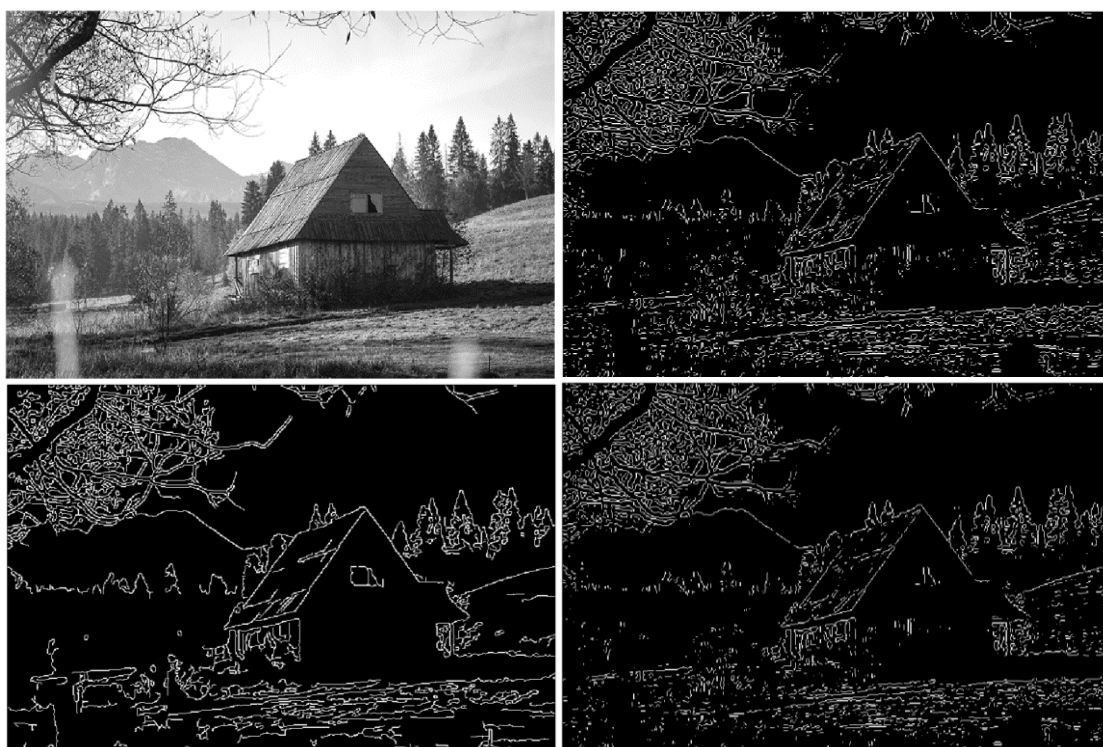
konvoluční masky patří Robertsův operátor (maska o velikosti 2x2, Obrázek 16), Cannyho, Prewittové, Sobelův a Robinsonův operátor (maska o velikosti 3x3).

**b) Hledání průchodů druhých derivací nulou**

U skupiny, která je založena na hledání průchodů druhé derivace nulou (také zvaná zero-crossing, viz Obrázek 16) jsou předními zástupci operátory Laplacián a Marr-Hildreth, jež nepodléhají změně v závislosti na rotaci jádra, a tudíž mohou být vypočteny jedinou konvoluční maskou.

**c) Lokální aproximace obrazové funkce parametrickým modelem**

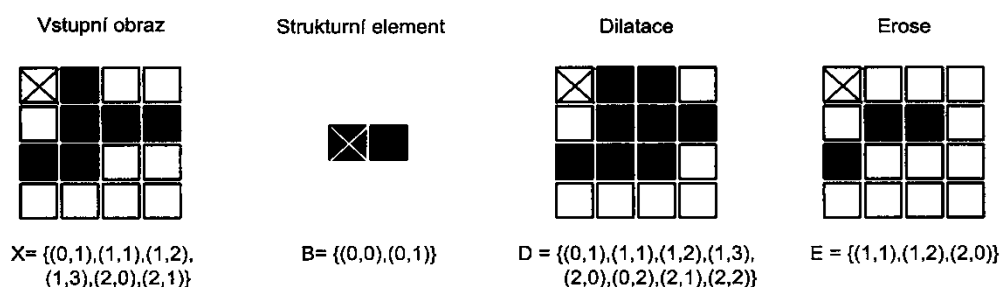
Princip této metody je založen na porovnání jednoduchým parametrickým modelem hran, kterým může být například polynom dvou proměnných.



Obrázek 16: Detekce hran za pomoci tří různých operátorů. Zleva původní snímek, Robertsův operátor, Cannyho operátor (vlevo dole), Zero-crossing (vpravo dole), (vlastní)

- **Morfologické operace** – technika nelineárního zpracování geometrických struktur obsahující rozhodování. Základními morfologickými operacemi jsou dilatace a eroze. Jejich vzájemnou kombinací pak získáváme operace tzv. morfologické otevření a uzavření. U těchto operací využíváme vždy vstupní obraz a strukturní element, jejichž počátky jsou označeny křížkem (Obrázek 17). Samotná realizace morfologické transformace (dilatace nebo eroze) odpovídá systematickému posouvání strukturního elementu po obraze. Výsledkem dilatace je, u binárního obrazu, vždy zvětšený objekt oproti objektu v obraze původním. Opačný jev představuje výsledek aplikování morfologické operace eroze, kdy je původní objekt zmenšený, a to proto, že části objektu, které jsou menší než velikost strukturního elementu, jsou odstraněny. Za

pomocí těchto operací je také možné vytvořit skelet objektu, který se často využívá u sledování objektů v obraze.



Obrázek 17: Základní binární morfologické operace – dilatace a eroze [1]

## 2.3 Segmentace obrazu

Jedním z nejnáročnějších kroků při zpracování obrazu je jednoznačně segmentace obrazových dat.

Segmentací digitálního obrazu se rozumí komplexní skupina metod postavených na různých principech, která slouží k automatickému rozdělení snímku na oblasti se společnými vlastnostmi, které poskytují pozorovateli potřebné informace.

Principem segmentace je rozklad obrazu na jednotlivé části, které jsou pro další zpracování důležité. Výsledkem by mělo být rozložení vstupního snímku do tzv. podobrazů, kde jednotlivé vrstvy nesou buď důležité, nebo méně podstatné informace. Tyto informace dále označujeme jako popředí (nesoucí důležitá data) a pozadí (nepotřebná data). Pro další analýzu se používají pouze oblasti zájmu.

Segmentaci rozdělujeme podle výstupního počtu oblastí na binární a multiregionální.

### Binární segmentace

Výstupem binární segmentace je obraz, který nabývá pouze dvou hodnot, nejčastěji 0 (černá) a 255 (bílá). Typickým příkladem této segmentace je segmentace metodou prahování, kdy pixelům s vyšší intenzitou jasu než je zvolený práh přiřadíme jednu z výše zmíněných hodnot. Stejně tak pro pixely s hodnotou menší než stanovený práh, přiřadíme zbylou hodnotu. Ukázka binární segmentace je znázorněna u statické metody segmentace prahováním (Obrázek 18 a Obrázek 19).

### Multiregionální segmentace

Multiregionální segmentací se rozumí obraz s více než dvěmi výstupními hodnotami v obraze. U této metody je možné si zvolit, kolika jasových hladin bude výsledný obraz nabývat. Ukázka multiregionální segmentace je taktéž znázorněna obrázkem a jeho histogramem s vyznačenými prahy segmentace (Obrázek 18 a Obrázek 19).

Jak už bylo zmíněno, segmentace je skupinou metod fungujících na různých principech, proto je potřeba si tyto metody rozdělit. Základními metodami segmentace obrazu jsou [12]:

- statické metody,
- metody na základě regionů v obraze,
- metody založené na detekci hran,
- znalostní metody,
- hybridní metody

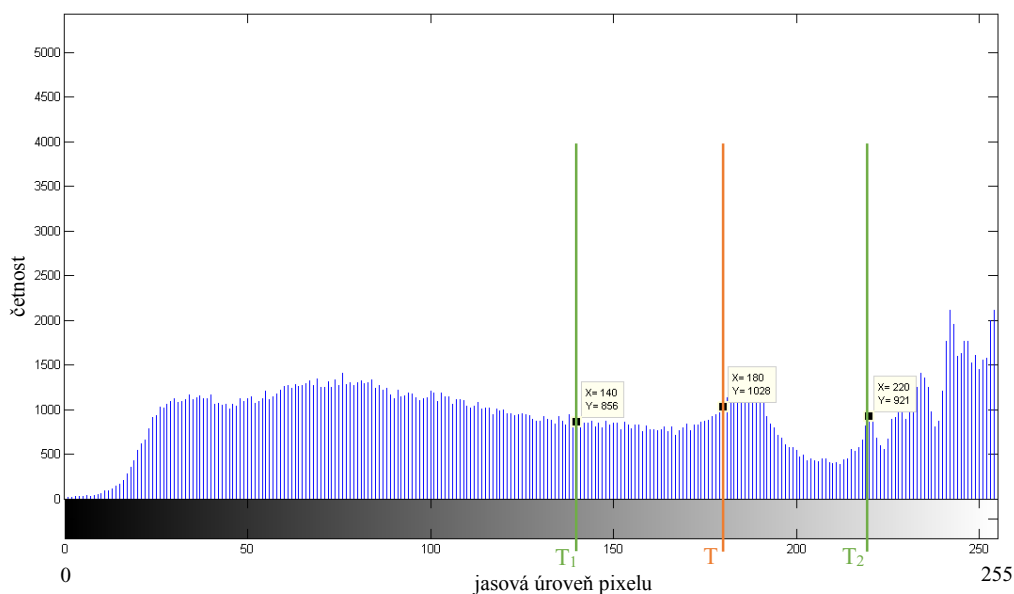
### Statické metody – metoda prahování

Metoda prahování, nebo také nazývána jako thresholding, je operace, při níž je obraz rozdělen do dvou (binární prahování) nebo více oblastí na základě daného prahu. Výsledkem je tedy oddělení oblasti zájmu od pozadí, v případě binárního prahování, viz Obrázek 18, se jedná o binární obraz, reprezentován následujícím zápisem (2.8), kde  $g(x,y)$  je výstupní obraz,  $f(x,y)$  je vstupní obraz,  $T$  - hodnota prahu, 0 - reprezentující černou barvu a 255 barvu bílou. Hodnoty prahů jsou vyznačeny v histogramu na Obrázek 19. Multiregionální segmentace, při zvolení dvou prahových hodnot je znázorněna taktéž na následujícím snímku i histogramu.

$$g(x,y) = \begin{cases} 0 & ; f(x,y) < T \\ 255 & ; f(x,y) \geq T \end{cases} \quad (2.8)$$



Obrázek 18: Znázornění segmentace metodou prahování. Zleva původní snímek, binární obraz s prahem  $T = 180$  (histogram oranžová značka), multiregionální obraz s prahy  $T_1 = 140$  a  $T_2 = 220$  (histogram zelená značka)



Obrázek 19: Histogram s vyznačenými prahy segmentace pro binární a multiregionální segmentaci

Nejčastějším problémem při aplikování této metody je zvolení hodnoty prahu. Pokud bychom chtěli segmentaci plně automatizovat, museli bychom zajistit velmi podobné vstupní snímky, co se týče expozice resp. jejich histogramů. Dále pak je možnost využití metody adaptivního prahování, kdy na základě zjištění histogramu vstupního snímku je vypočtena hodnota prahu nebo se mohou uplatňovat pravděpodobnostní přístupy.

### **Metoda na základě regionů**

Metoda na základě regionů, z ang. Region-based, detekuje oblasti v obraze na základě homogenních oblastí reprezentovaných barvou, intenzitou nebo texturou. Rozlišují se zde tři základní principy. Princip regionálního růstu, princip štěpení oblastí a kombinace těchto dvou principů je nazývána jako metoda štěpení a spojování oblastí (split and merge).

Nejjednodušší metodou je princip regionálního růstu, kde jsou zvoleny dané pixely v obraze a od nich se postupně oblasti rozšiřují až po celé segmenty. Zvolenému pixelu se říká iniciální bod neboli semínko. Výsledný obraz po segmentaci je tedy značně ovlivněn rozmístěním iniciálních bodů a na kritériích homogenity.

Princip štěpení oblastí funguje na podobné bázi jako princip regionálního růstu, s tím rozdílem, že začíná jako jedna velká oblast, která se na základě kritérií homogenity dělí na menší části. Bylo zjištěno, že i přes stejná kritéria homogenity dochází k jiným výsledkům segmentace [1].

### **Metoda založená na detekci hran**

Tato metoda segmentace využívá již zmíněné detekce hran pomocí různých hranových operátorů (kapitola 2.2 – detekce hran). Výsledkem této detekce nedostáváme úplně přesné informace a je potřeba takto vytvořený obraz dále zpracovat. Hlavním důvodem je špatné detekování hran v místech s přítomným šumem, a nebo jejich neúplná detekce. Jedná se tedy o další metody rekonstruující obrazové hrany a jednou z nich je například metoda štěpení a spojování oblastí [1].



### 3 Metody detekce obličeje z obrazu

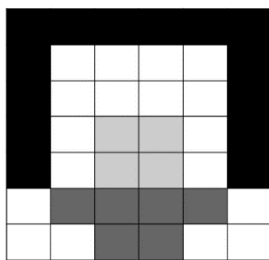
Vzhledem k velkému rozmachu digitální techniky pro rozmanité účely, narůstají požadavky na zabezpečení různých systémů a aplikací, a obličej je jedním z předních biometrických rysů, které používáme k identifikaci osob každý den. Díky tomu, že každý z nás je jedinečný je možné přizpůsobit počítačové vidění k požadovanému účelu. Na základě zpracování obrazových dat můžeme velice spolehlivě rozpoznávat osoby podle jejich obličeje, s výjimkou jednovaječných dvojčat nebo dvojníků. Pro základní rozpoznávání obličejů postačí klasické 2D snímky pořízené fotoaparátem. V případě vyšších nároků na spolehlivost a bezpečnost systému je možné využít rozpoznávání z 3D snímků spolu s termografickými snímky nebo také v kombinaci s různými dalšími biometrickými znaky, jako např. rozpoznání oční duhovky, sítnice a otisků prstů. Tyto metody jsou však finančně značně nákladnější. Abychom přiměli počítač rozpoznávat osoby, musíme jej prvně naučit obličej v obraze detekovat.

Metod pro detekci obličeje z obrazu existuje velké množství, avšak nejsou pevně kategorizovány. Přesto je lze rozdělit do několika základních skupin [13].

#### 3.1 Metody na základě znalostí o lidské tváři – Knowledge-based methods

U metody o znalosti lidské tváře se využívá známých charakteristických rysů obličeje a vztahů mezi nimi. Mezi nejčastěji používané charakteristiky patří pozice očí, úst a případně i nosu. Stanovení těchto pravidel je velmi složité, a to proto, že pokud jsou zvolena pouze základní kritéria, program může označit za obličej i předměty, které obličejem nejsou. Naopak je tomu v případě, že jsou kritéria stanovena příliš detailně a detektor nemusí obličej v obraze vůbec najít. Proto je vhodné volit určitý kompromis společně s kaskádním řazením úrovní daných kritérií, kterými vstupní obraz prochází.

Mezi nejčastěji využívanou metodu na základě znalostí patří hierarchická metoda se třemi úrovněmi rozhodování. Výhodou této metody je její efektivita, a to především v případech, kdy snímáme obličej v různých vzdálenostech od senzoru, tzn. že velikost obličeje se ve snímku značně liší. V první fázi rozhodovacího procesu je vytvořena hrubá mozaika, reprezentující oblasti lidské tváře v závislosti na rozložení jasu jednotlivých obrazových bloků (Obrázek 20). Oblastmi lidské tváře se rozumí především oči, nos, rty, vlasy atd.



Obrázek 20: Možnost vytvoření obrazových bloků reprezentující obličej [14]

Zpracovávaný obraz (Obrázek 21) je tedy převeden do nižšího rozlišení a porovnáván s výše uvedenou mozaikovitou maskou. Jednotlivé elementární prvky mozaiky reprezentují průměrnou hodnotu jasu

okolních pixelů. Nezáleží pouze na hodnotách jasů, ale také na vzájemném rozdílu (kontrastu) sousedících oblastí.



*Obrázek 21: Postupné snižování rozlišení zkoumaného snímku [13]*

Touto metodou jsou vybrány snímky, ve kterých se potenciálně mohou vyskytovat lidské obličeje.

V druhém kroku se využívá lokální ekvalizace histogramu, sloužící pro zlepšení kontrastu oblasti a tím k přesnější detekci obrazových hran (viz kapitola 2.2). Po detekci hran je tedy získán přibližný model obličeje, který má daný poměr své šířky a výšky.

Ve třetí fázi se již systém zaměřuje na jednotlivé oblasti lidské tváře. Nejprve se detekuje pozice očí a obočí tak, že se v horní oblasti modelu obličeje hledají lokální maxima jasu (odpovídající oblasti mezi očima a obočím) v daném rozsahu vzdáleností vycházejících z poměru tváře. Od těchto maxim je následně hledáno lokální minimum, které představuje pozici očí. Detekce nosu a úst je v této chvíli snadnou záležitostí, kdy se detekce zaměřuje na oblasti mezi očima směrem dolů. Špička nosu je reprezentována lokálním maximem a nosní dírkou lokálními minimy. Ústa jsou reprezentována homogenními oblastmi nacházejícími se několik řádků pod nosem. Pokud systém detekuje veškeré tyto oblasti, pak je dané místo v obraze označeno jako obličej.

### **3.2 Metody na základě neměnných rysů obličeje – Feature invariant methods**

Metody na základě neměnných rysů obličeje využívají především takové rysy obličeje, které jsou nezávislé na úhlu natočení hlavy vzhledem ke snímáči. Jedná se především o detekci obličeje podle barvy kůže, která s sebou nese jistá omezení. Nejzásadnějším problémem je různá barva kůže především u různých etnických skupin. Odstín barvy kůže také ovlivňují vnější světelné podmínky [13].

Studie prokázaly, že i přes všechny tyto omezující faktory je tato metoda efektivní. Důraz této metody je kladen na detekci pixelů v určitém rozsahu barev daného barevného modelu (nejčastěji se jedná o modely RGB, YCbCr nebo HSV). Jedná se tedy o segmentaci obrazu na základě barvy kůže. Rozpoznáním pixelů, v daném barevném rozsahu, detekce nekončí. Je nutné rozlišit pixely, které odpovídají obličejům a pixely, které jsou pozadím snímku. Proto je nutné, spolu s touto metodou, využít i další metody, které zajistí správné rozpoznání lidské tváře. Lze říci, že tato metoda relativně spolehlivě a efektivně oddělí místa v obraze, ve kterých se může nacházet obličej a zavrhnout oblasti, které jsou pozadím snímku. Na základě barevného spektra lze detekovat nejen barvu kůže, ale také pozici očí a úst [15] [17].

Mezi nejefektivnější metody detekce lidské kůže patří zejména detekce z barevného prostoru YCbCr (kapitola 2.1). Tento barevný prostor není tak moc náchylný na změnu osvětlení scény jako model RGB. I přesto vzhledem k různým etnickým skupinám je spektrum barev odpovídajících lidské kůži docela široké, avšak oproti RGB velmi přijatelné. Jako první je potřeba převést barevný model RGB na model YCbCr (kapitola 2.1). Následně je každý pixel obrazu testován, zda spadá do intervalu hodnot patřící

lidské kůži. Zastoupení hodnot jednotlivých složek YCbCr pro lidskou kůži bylo experimentálně zjištěno a je znázorněno v následujících intervalech (3.1) [17] [18].

$$\begin{aligned} 110 < Y < 190 \\ 77 < Cb < 127 \\ 137 < Cr < 177 \end{aligned} \quad (3.1)$$

Pokud hodnota pixelu odpovídá hodnotě v daném rozsahu, je tento bod označen za pixel, který může patřit lidské kůži. Tímto způsobem je získán binární obraz, kde bílá barva reprezentuje barvu kůže (Obrázek 22 uprostřed). Následně je využito morfologických operací (popsáno v kapitole 2.2) sloužících k odstranění detekovaných oblastí s malými rozměry určenými předem zvoleným prahem, které nemohou představovat obličej (Obrázek 22 vpravo). Tyto oblasti mohou vznikat detekcí kůže na jiných místech těla nebo přítomností šumu v obraze. Hodnota prahu musí být experimentálně zvolena tak, aby respektovala vlastnosti vstupního obrazu a snímané scény.



*Obrázek 22: Původní snímek, binární obraz detekce kůže z modelu YCbCr, binární snímek po použití morfologických operací [17]*

Samotná detekce tohoto rozsahu barev však není finálním řešením, vzhledem k tomu, že podobná barva se může vyskytovat v pozadí nebo při špatně zvoleném osvětlení, kdy do této škály barev mohou spadat také vlasy (převážně světle hnědé). Často se tedy využívá ořez dané oblasti na základě poměru šířky a výšky obličeje, který často bývá v rozmezí 0,8 – 0,85. Dalším ovlivňujícím faktorem je volba oblečení (viz předcházející obrázek), kdy osoba může mít vysoký rolák nebo naopak hluboký výstřih a tudíž je detekována jiná plocha kůže a vzniká problém při výběru dané oblasti na základě výše zmíněného poměru šířky a výšky obličeje.

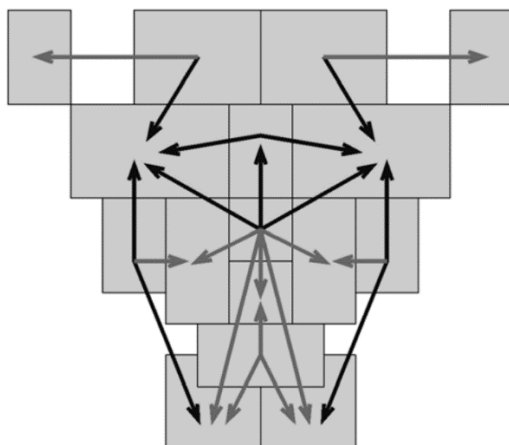
Detekce obličeje na principu barvy kůže je dále doprovázena detekcí hran z monochromatického obrazu vrstvy Y a také hledáním obličejových znaků jako jsou oči, nos a ústa. Na tuto metodu často navazuje korelace oblasti obličeje s referenčním snímkem v databázi. V této chvíli se různé metody pro detekci obličeje začínají kombinovat a prolínat.

### **3.3 Metody na základě porovnávání se vzorem – Template matching**

U této metody se využívá porovnávání vzorů nebo také šablon vůči vstupnímu snímku. Detekce je založena na výpočtu korelačního koeficientu vstupního snímku obličeje resp. pouze detekované oblasti kůže spolu se vzorovým modelem obličeje. Složitost těchto metod je spjata především s tvorbou referenčních šablon, jejichž tvorba je časově náročná a složitá. Tato metoda se dá použít také pouze pro porovnávání jednotlivých částí obličeje, jako jsou například oči nebo ústa. Nevýhodou těchto metod je velká citlivost na úhel natočení hlavy nebo vzdálenosti od snímáče. Tyto problémy však lze řešit různými deformovatelnými vzory, ale opět se zde stupňuje celková složitost a náročnost systému [13].

#### **Předem stanovené referenční vzory**

Metoda detekce obličeje na základě porovnávání se vzorem se začala vyvíjet na šablonách reprezentujících obličej z pohledu zepředu. Tento referenční vzor se skládal z několika menších tzv. subšablon, které sloužily pro detekci očí, tváří, nosu, úst, obrysu obličeje atd. Jednotlivé subšablony jsou znázorněny šedými oblastmi (Obrázek 23). Vzájemné vztahy mezi těmito oblastmi jsou znázorněny šipkami v téže obrázku. Vztahy, které jsou označeny černou šipkou, se nazývají jako nezbytné vztahy (essential relations). Šedé šipky se označují jako potvrzující vztahy (confirming relations).



*Obrázek 23: Referenční vzor obličeje skládající se z 16 subšablon a 23 vzájemných vztahů (11 nezbytných a 12 potvrzujících) [13]*

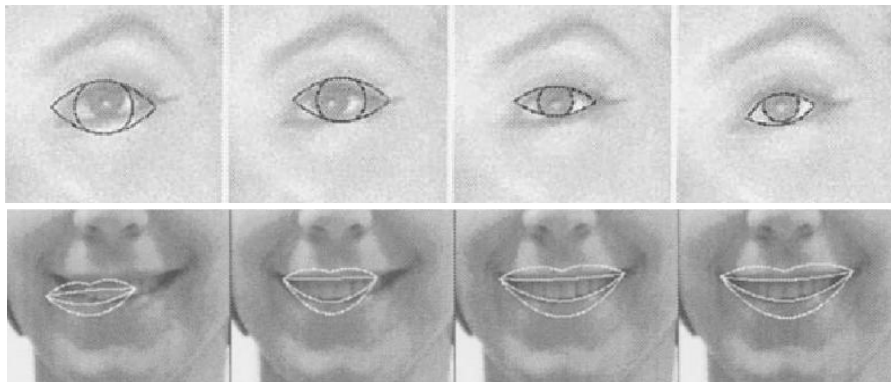
Porovnání této šablony a vstupního snímku se nejčastěji provádí metodou vzájemné korelace. Pro zvýšení přesnosti detekce obličeje se často na vstupní snímek nejprve aplikuje některý z hranových detektorů, vybrané geometrické transformace a až pak se provádí samotná korelace [13].

Jako referenční vzory lze také využívat obrazové siluety získané pomocí PCA (Principal Component Analysis), což je analýza hlavních komponent obličeje, které jsou zastoupeny bitovým polem. Na takto vzniklé sadě siluet se kombinuje využití základních technik zpracování obrazu (detekce hran, prahování, morfologické operace) spolu se zobecněnou Houghovou transformací, pomocí které je možné identifikovat libovolné tvary v obraze [19].

### **Deformovatelné referenční vzory**

Vzhledem k jistým omezením předem stanovených referenčních vzorů se začíná stále častěji využívat šablon, které se dokáží přizpůsobovat reálným znakům lidského obličeje. Šablony s takovými vlastnostmi jsou označovány jako deformovatelné.

Deformovatelné vzory jsou definovány parametrickými přímkami a plochou. Šablona se postupně posouvá po obraze až do chvíle, kdy narazí na potencionální místo hledaného znaku, např. oka nebo rtů. Poté se začne deformovat (smršťovat, roztahovat nebo otáčet) v závislosti na reálných hranách obličejového znaku, viz Obrázek 24.



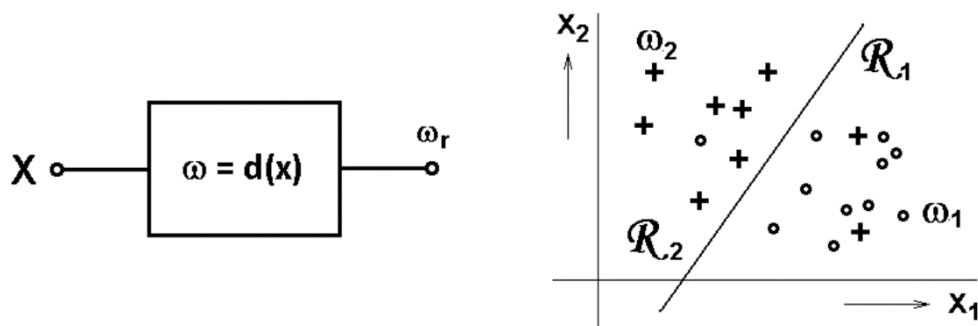
Obrázek 24: Ukázka deformovatelných referenčních vzorů oka a úst

### 3.4 Metody na základě vzhledu – Appearance-based methods

V těchto metodách se využívá detekce obličeje na základě strojového učení. Jádrem celé metody je klasifikační algoritmus, kterému jsou předkládány výřezy snímků s lidskými obličeji nebo bez nich. Algoritmus musí umět vyhodnotit, zda je ve výrezu lidská tvář obsažena, či nikoli. Následně si detektor poznačí ta místa, která byla označena pozitivně. Jedná se tedy o metodu trénování, na základě trénovacích dat.

#### Klasifikace dat

Klasifikací rozumíme separaci množiny abstraktních objektů v obraze na konečný počet dílčích skupin, které nazýváme klasifikačními třídami nebo také shluky. Tyto třídy obsahují data, jež mají podobné charakteristické vlastnosti a jejichž příznaky jsou téměř neměnné. Systém (již zmíněný klasifikační algoritmus nebo také klasifikátor) pro rozdělování jednotlivých objektů do tříd se řídí tzv. klasifikačními kritérii,  $\omega = d(X)$  viz Obrázek 25, kde  $d(X)$  je funkce argumentu  $X$  reprezentujícího vstupní data. Výstupem klasifikátoru je tedy hodnota  $\omega_r$ , která představuje danou klasifikační třídu, do níž byl vstupní objekt zařazen, viz klasifikační třídy v následujícím obrázku. U rozpoznání obličeje se jedná o klasifikaci multitřídní tzn., že se předpokládá, že databáze obsahuje minimálně 2 vzory obličejů.



Obrázek 25: Klasifikátor (vlevo) a klasifikační třídy  $\omega_1$  a  $\omega_2$  (vpravo) [22]

Klasifikace dat založená na základě strojového učení pochází ze dvou možných přístupů: supervizorního a nesupervizorního učení.

### **Učení bez učitele – unsupervised learning**

Učení bez učitele představuje takový systém, který má k dispozici pouze vstupní data, nikoli informace o jejich správném zařazení do tříd nebo dokonce počtu těchto tříd. V případě, že trénovací množina dat u tohoto přístupu učení není k dispozici, pak má klasifikátor dostupný pouze obecný tvar rozhodovacího procesu a nastavování jeho parametrů probíhá současně s klasifikací. V tomto případě musí systém nejdříve v parametrickém prostoru odhalit třídy objektů. Například pomocí vyhledávání shluků v parametrickém prostoru. Typickým příkladem učení bez učitele je shluková analýza nebo hierarchické shlukování a často se tento přístup používá v automatizovaných zařízeních nebo i pro screening známých procesů, které nevykazují očekávané výsledky [22] [24].

### **Učení s učitelem – supervised learning**

Učení s učitelem představuje apriorní informaci popisující inicializační klasifikaci dat do předurčeného počtu klasifikačních množin, tzn. že pro vstupní data je předem určeno správné zařazení v klasifikačních třídách. Děje se to na základě předem stanovené trénovací neboli učební sady dat, která obsahuje jak vstupní snímky (features), tak i informaci o správnosti jejich zařazení do tříd (class). Nejjednodušším představitelem této metody je klasifikátor k-NN (Nearest Neighbour) využívající podobnost nejbližších sousedů na základě stanovené metriky, např. Euklidovská vzdálenost [22] [23].

Nevýhodou této metody je nízká flexibilita a vysoké náklady (související s tvorbou množiny vstupních údajů).

Vzhledem k míře správnosti klasifikace této trénovací sady se algoritmy dále rozdělují na učení s dokonalým či nedokonalým učitelem. Existuje také kombinace těchto dvou přístupů, která se označuje jako semi-supervised learning [22] [24].

## 4 Metody sledování a rozpoznání obličejů z obrazu

Nejvýznamnějším přínosem, při rozpoznávání obličejů, je bezesporu informace nesoucí přesnou totožnost jedince na snímku. Metod a postupů pro získání této informace je v současnosti mnoho a liší se především svou složitostí a přesností. V závislosti na vstupních obrazových datech rozlišujeme dvě kategorie detekce a rozpoznávání identity. Jedná se tedy o statické obrazy (klasické 2D snímky) a dnes především o dynamické obrazy, jiným slovem videosekvence (2D snímky s danou časovou posloupností). Nejdokonalejší verzí těchto systémů je sledování a rozpoznávání tváří reálném čase (realtime face tracking and recognition).

### 4.1 Face tracking - sledování detekované tváře

Face tracking neboli sledování lidské tváře navazuje na již zmíněnou detekci obličejů (viz kapitola 3). Jedná se o počítačovou techniku, jejíž cílem je udržet stopu (měnící se pozici) detekované tváře v obraze v rámci videosekvence. Nejedná se vůbec o snadný úkol, protože se v průběhu času může měnit osvětlení, podoba a vzhled tváře, v důsledku pohybu, natočení, mimiky atd. Tyto změny jsou důvodem, proč nemůžeme chápat sledování obličejů jako nepřetržitou detekci tváře.

Náročnost sledování obličejů ve videu vyplývá ze dvou hlavních důvodů, které se vzájemně ovlivňují. Prvním z nich je rozlišení a celková kvalita zkoumaného obrazu. Druhým problémem ovlivňujícím možnost sledování tváře je výpočetní náročnost algoritmu. Tzn., čím kvalitnější vstupní data máme, tím přesněji jsme schopni tvář v obraze sledovat na úkor delšího časového zpracování. Zde se tedy objevuje problém při sledování obličejů v reálném čase a je potřeba volit vhodný kompromis mezi kvalitou a rychlostí zpracování (jak snížením kvality obrazu, tak i vhodnou metodou a optimalizací sledovacího algoritmu).

V současnosti existuje mnoho metod hledajících neoptimálnější řešení v oblasti sledování obličejů. Metody se rozdělují do několika skupin na základě jejich vlastností. Rozlišují se algoritmy, které pracují na bázi online nebo offline učení (online-learning, offline-learning), algoritmy rozpoznávající jeden nebo více objektů v obraze (single-object, multi-object), nebo také metody, které se liší podle principu detekce objektu ve videosekvenci, mezi které patří algoritmy generativní a diskriminativní.

Další dělení těchto metod je založeno na základě reprezentace dat, jejich zpracování, odhadu pohybu objektu v obraze a modelování vzhledu snímané scény. Mezi dvě základní metody patří algoritmus optického toku (Optical Flow) a algoritmus kontinuálního adaptivního posunu Mean Shift (CAMShift) [32], [33].

#### 4.1.1 Optical Flow

Sledovací algoritmus optical flow je definován jako změna jasu pixelů ve videosekvenci. Pro sledování obličejů přes posloupnost snímků je využíván algoritmus optického toku Lucas-Kanade-Tomasi (KLT), který má za úkol odhadovat pohyb extrahovaných rysů obličejů v sousedních snímcích videa. Je možné stanovit druh a rychlost pohybu každého pixelu v obraze, jinými slovy jejich posunutí, které představuje vzdálenost, o kterou se pixel posunul mezi předchozím snímkem a aktuálním snímkem. Princip funkce algoritmu KLT bude podrobněji popsán v kapitole 5.3.3.

Optický tok se rozlišuje na hustý a řídký. Výpočet hustého optického toku je dosti obtížný úkol, protože tento výpočet musí vzít v úvahu každý jednotlivý pixel obrazu. Možným řešením je použít interpolaci na pixely, které je obtížné sledovat, pomocí informací z těch pixelů, které lze dobře sledovat. Toto řešení však vede k vyšší výpočetní náročnosti. Naopak řídký optický tok spoléhá na určení podmnožiny bodů,



kteřé mají být sledovány. Tato podmnožina je složena z pixelů, které lze dobře sledovat (z ang. „good features to track“), jako jsou například hrany detekované hranovými detektory.

I když u této metody je obtížnější pracovat v reálném čase s hustým optickým tokem kvůli jeho výpočetním nárokům, v některých souvislostech je větší zájem o získání vynikající vizuální kvality než o rychlý výkon, to se uplatňuje například ve filmové postprodukci [33].

#### 4.1.2 Sledování CAMShift

Algoritmus CAMShift (Continuously Adaptive Mean Shift) je přizpůsobením algoritmu Mean Shift pro sledování objektů. Algoritmus Mean Shift je robustní neparametrická technika, která zjišťuje gradient rozdělení pravděpodobnosti za účelem nalezení vrcholu distribuce, ze kterého se pak vychází v následujícím snímku videosekvence.

Primární vylepšení CAMShift spočívá v tom, že tato metoda využívá nepřetržitě adaptivní rozložení pravděpodobnosti (tj. distribuce, které mohou být přepočítány pro každý snímek), zatímco metoda Mean Shift je založena na statických rozloženích pravděpodobnosti, které se neaktualizují, pokud nejsou zaznamenány významné změny ve tvaru, velikosti nebo barvě sledovaného objektu.

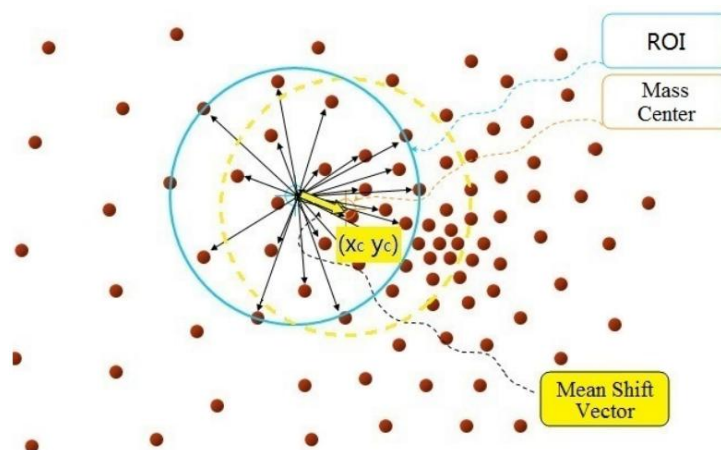
Algoritmus CAMShift pracuje v následujících krocích:

1. Nastavení oblasti zájmu ROI (z ang. Region of Interest) na celý obraz.
2. Výběr počátečního umístění okna pro vyhledávání metodou Mean Shift. Vybraná pozice představuje cílové rozložení hodnot pixelů, které mají být sledovány.
3. Výpočet distribuce pravděpodobnosti barev v ROI vycentrovaném vyhledávacím oknem Mean Shift.
4. Iterace algoritmu Mean Shift pro nalezení těžiště objektu. Uložení nultého neboli aktuálního okamžiku (distribuční oblast) a lokace těžiště (Mass Center).
5. Pro následující snímek je vycentrováno vyhledávací okno na místo mediánu nalezeného v kroku 4 a nastavení velikosti okna na funkci nultého momentu. Poté následuje vrácení se ke kroku 3 a proces se opakuje.

Vytvoření pravděpodobnostní distribuční funkce odpovídá krokům 1 až 3. Pro inicializaci ROI používáme oblast obličeje detekovanou na posledním snímku sekvence. Poté je potřeba vypočítat barevný histogram odpovídající této oblasti v barevném prostoru HSV (Hue Saturation Value). Tento barevný prostor je podrobněji popsán v [34]. Vzhledem k situaci, že lidské barvy kůže mají malý rozdíl v kanálech S (sytnost) a V (jas), používá se pro výpočet pouze H (odstín) kanál. Využitím pouze jednoho kanálu se zkracuje výpočetní náročnost algoritmu. Histogram je následně kvantován do zásobníků, které slouží pro seskupení pixelů s podobnými barevnými hodnotami. Poté se použije zpětná projekce histogramu. Jedná se o jednoduchou operaci, která spojuje hodnoty obrazových bodů v obraze s příslušnými hodnotami v zásobníku. To znamená, že se určí barvy v obraze, které spadají do ROI a za pomoci zpětné projekce jsou tyto pixely v obraze vyobrazeny.

Tento algoritmus tedy neustále přepočítává nové hodnoty vektoru  $(x_c, y_c)$  pro oblast ROI, vypočítanou z předchozího snímku, do doby kdy dojde k nevýznamnému posunu ROI v obraze, tzv. konvergenci. Obrázek 26 ukazuje jednu iteraci výše popsaného algoritmu Mean Shift.



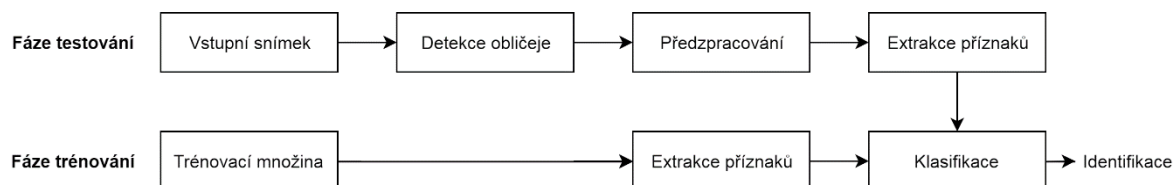


Obrázek 26: Znáznornění jedné iterace algoritmu Mean Shift [32]

## 4.2 Rozpoznání tváře

Rozpoznávání tváří je proces, který porovnává obličej vstupního snímku se všemi obličejí v databázi a určuje identitu člověka. Identifikace testovaného obrazu se provádí lokalizací snímku v databázi, která má s testovaným obrazem nejvyšší podobnost.

Obecným postupem při rozpoznávání obličejů se rozumí pořízení snímku, ověření a detekce obličeje v obraze, viz kapitola 3.4. Následně je detekovaný obličej ze vstupního obrazu ještě předzpracován podle potřeb klasifikátoru, jedná se především o zarovnání detekovaného obličeje podle očí na stanovenou pozici, změnu měřítka nebo převodem snímku na monochromatický. V druhé fázi je potřeba natrénovat klasifikační algoritmus za pomoci extrahovaných příznaků z trénovací množiny dat. Poté jsou porovnány příznaky vstupního snímku se snímky v databázi klasifikátorem a na základě skóre podobnosti je vyhodnocena identita člověka (Obrázek 27).



Obrázek 27: Obecný postup rozpoznávání obličejů (vlastní)

### Extrakce příznaků

Extrakci příznaků se rozumí vytažení důležitých vlastností obličeje z obrazu podle zvolené metody rozpoznání obličeje. Výsledkem tohoto procesu je specifický vektor příznaků pro zkoumaný obličej, nacházející se v prostoru příznaků. Důležitým předpokladem je balance počtu vektorů na danou třídu, kdy tyto vlastnosti ovlivňují výkonnost klasifikátoru. V praxi jsou tyto prostory příliš velké, aby umožnily rychlé a zároveň robustní rozpoznávání objektů. Běžným způsobem, jak se pokusit tento problém vyřešit, je využití technik pro redukci rozměrů dimenzí, jako je například statistická metoda redukce příznaků PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis) a další [21].

Metody pro identifikaci osob podle obličeje lze rozdělit do dvou základních kategorií dle přístupu na geometrické a holistické.

#### a) Geometrické metody

Geometrickými metodami (označováno také jako strukturální) se rozumí takové metody, které jsou založené na dominantních obličejových znacích, které jsou charakteristické svými rozměry a vzájemnou pozicí vůči sobě navzájem. Jedná se především o vlastnosti obličeje jako je obočí, oči, nos a ústa. Úkolem geometrických metod je tyto vlastnosti analyzovat, normalizovat a porovnat s databází obrazů. Výhodou těchto metod je relativně malá závislost na osvětlení [25]. Hlavními představiteli geometrických metod jsou metody shlukových grafů, metody založené na porovnání obličejových vektorů nebo ASM (Active Shape Model), AAM (Active Appearance Model) [1] [26].

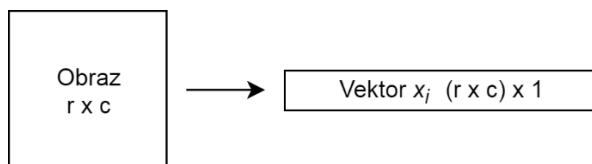
#### b) Holistické metody

Holistické metody jsou oproti geometrickým metodám založeny na obličeji jako celku (nikoli pouze jeho vlastností), a tudíž nejsou při rozpoznávání tváře zanedbány žádné jeho části. Tohle je pravděpodobně největší výhodou těchto metod, protože porovnáváním globální reprezentace obličeje se zvyšuje míra správnosti identifikace, která je dána statistickým vyhodnocením relativní pravděpodobnosti. Na druhou stranu, jeli porovnáván obličej jako celek není možné docílit stejné rychlosti zpracování, jako když jsou porovnávány jen některé vybrané oblasti. V současnosti však existují techniky, které jsou schopny výpočetní čas značně snížit. Mezi základní holistické metody se řadí již zmíněné statistické metody PCA (v literatuře označována jako Eigenfaces) a LDA (Fisherfaces), nebo také metody SVM (Support Vector Machine), LBPH (Local Binary Patterns Histogram), neuronové sítě a další [21] [27].

### 4.2.1 Eigenfaces

Eigenfaces je statistická technika rozpoznání obličeje, která je obecně označována jako analýza hlavních komponent PCA (Principal Component Analysis) a jejím úkolem je transformovat prostor obličeje do nového prostoru, vytvořeného na základě rozložení vstupních dat. Výsledná projekce této metody má za úkol maximalizovat variaci dat mezi všemi obrazy (Obrázek 30). Eigenfaces představuje metodu strojového učení bez učitele, kdy vzniklý prostor nemá informaci o tom, do jaké třídy přísluší daný vstupní vektor (kapitola 3.4 – učení bez učitele). Snahou tohoto přístupu je najít důležité informace sloužící k rozpoznání obličejů v obraze a tím redukovat počet dimenzí s co nejmenší informační ztrátou. Tento prostor se skládá z jednotlivých os, seřazených na základě míry variace dat [1] [21].

Princip rozpoznávání založený na metodě PCA se skládá z několika kroků. Jako první je potřeba vytvořit z obrazů o daných rozměrech ( $r \times c$ ) řádkové vektory  $x_i$  o délce  $n$  (Obrázek 28). Pro příklad, máme-li obraz o velikosti 100 x 100 pixelů, výsledný vektor  $x_i$  bude obsahovat 10 000 prvků, přičemž každý tento prvek reprezentuje jasovou intenzitu pixelu v místě určeném svým indexem [1] [28].



Obrázek 28: Znáznornění převodu obrazu na řádkový vektor (vlastní)

Následně je zapotřebí spočítat střední hodnotu  $m_x$  všech bodů vektoru  $x_i$  (rovnice 4.1) a provést centrování bodů  $c_i$ , podle rovnice 4.2.

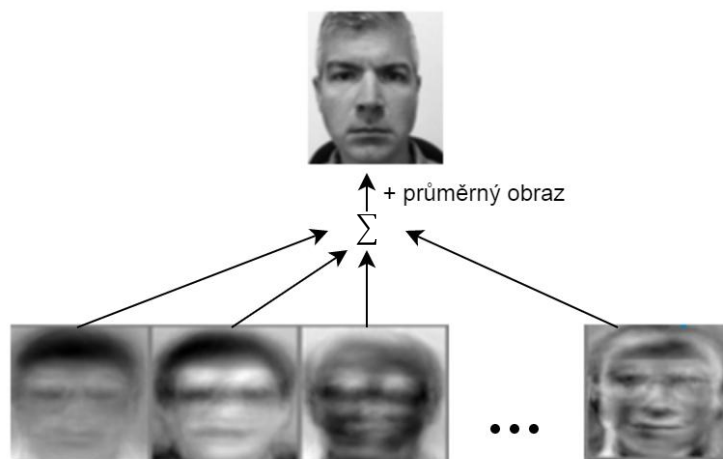
$$m_x = \frac{1}{n} \cdot \sum_{i=1}^n x_i \quad (4.1)$$

$$c_i = x_i - m_x \quad (4.2)$$

V dalším kroku je počítána kovarianční matice  $C$  podle vztahu 4.3.

$$C = \frac{1}{n} \cdot \sum_{i=1}^n c_i \cdot c_i^T \quad (4.3)$$

Takto vzniklá matice je pro běžné používání v praxi příliš velká a je potřeba ji zredukovat tak, aby bylo její využití efektivní a zároveň neztratila cenné informace. Cílem je určit vlastní vektory a vlastní hodnoty kovarianční matice  $C$  a seřadit je vzestupně podle jejich velikosti. Následně je proveden výběr prvních  $k$  vlastních vektorů podle toho, jak přispívají k popisu odchylek od průměrného obrázku (z tohoto důvodu mohou vlastní vektory připomínat obličejové a odtud vzniká název eigenfaces, viz Obrázek 29 dole). Vlastní vektory s nejmenšími hodnotami jsou zanedbány a generuje se množina dat v nové reprezentaci, která je již pro klasifikaci vhodná [1] [28].



*Obrázek 29: Zpětná transformace obrazu z redukovaného prostoru vlastních obličejů vzniklého metodou PCA (upraveno podle [29])*

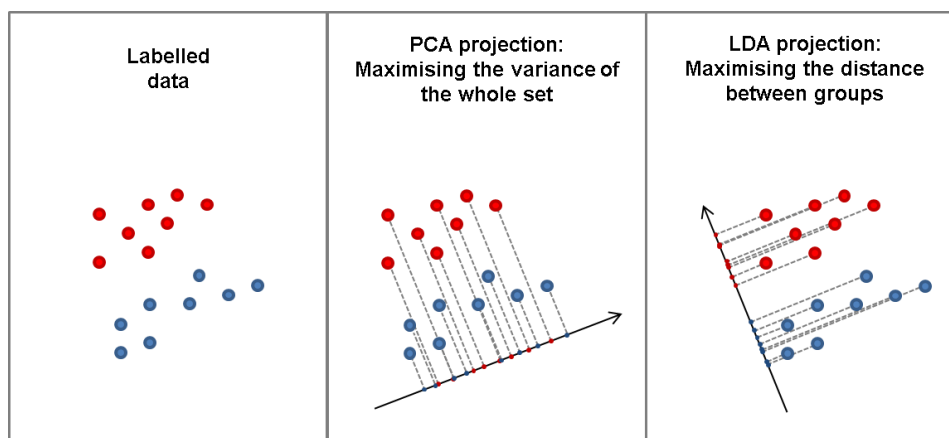
Na předcházejícím obrázku je znázorněna zpětná rekonstrukce obrazu do původního prostoru. Výsledný snímek (v obrázku nahoře) je oproti původnímu snímku méně ostrý, a to z důvodu, že je tento obraz vytvořen z nového, ale redukovaného prostoru. Při rozpoznávání obličejů se porovnávají vzdálenosti mezi vektory vlastních obličejů na základě stanovené metriky (Euklidova metrika, Čebyševova metrika a další). Vektor, který je podle dané metriky nejbližší k vektoru testovaného obličejové, určuje identitu jedince.

## 4.2.2 Fisherfaces

Fisherfaces je také statistickou metodou, která je ale založena na metodě LDA (Linear Discriminant Analysis). Princip této metody je podobný jako u PCA, kde cílem této metody je rozeznávat obličejové v prostoru příznaků vytvořeném lineární projekcí trénovacích obrazů. Snaha této metody také spočívá v redukci dimenzí prostoru pro hledání vzorů.

Na rozdíl od Eigenfaces je tato metoda založena na strojovém učení s učitelem, kdy jsou algoritmu spolu se vstupními obrazy předkládána data identifikující obličejové na snímcích. Nejdůležitějším úkolem u

metody LDA není zachytit nejpodstatnější informace obličeje, ale především vytvořit prostor tak, aby byly body reprezentující stejný obličej co nejlépe separovány od bodů ostatních obličejů a zároveň aby byla variace v rámci stejné třídy co nejmenší (Obrázek 30). Výhodou fisherfaces je lepší robustnost na změnu výrazu nebo na změnu směru osvětlení obličeje. Při strojovém učení s učitelem se také často objevuje jeden fenomén. Označuje se jako přeučení nebo také přílišná adaptace na trénovací sadu dat. To má za následek, že při výskytu nové tváře nelze předvídat, zda tato nová tvář bude v novém prostoru dostatečně odlišená od ostatních bodů patřících obličejům z trénovací sady [1] [30] [31].



Obrázek 30: Vymezení projekce prostoru a vlastních bodů pro dvě třídy metodou PCA a LDA [30]

Ukázka vlastních obrazů Fisherfaces, je znázorněna na obrázku (Obrázek 31).



Obrázek 31: Vlastní obličeje vytvořené metodou PCA (nahore) a LDA (dole) [31]

### 4.3 Neuronové sítě

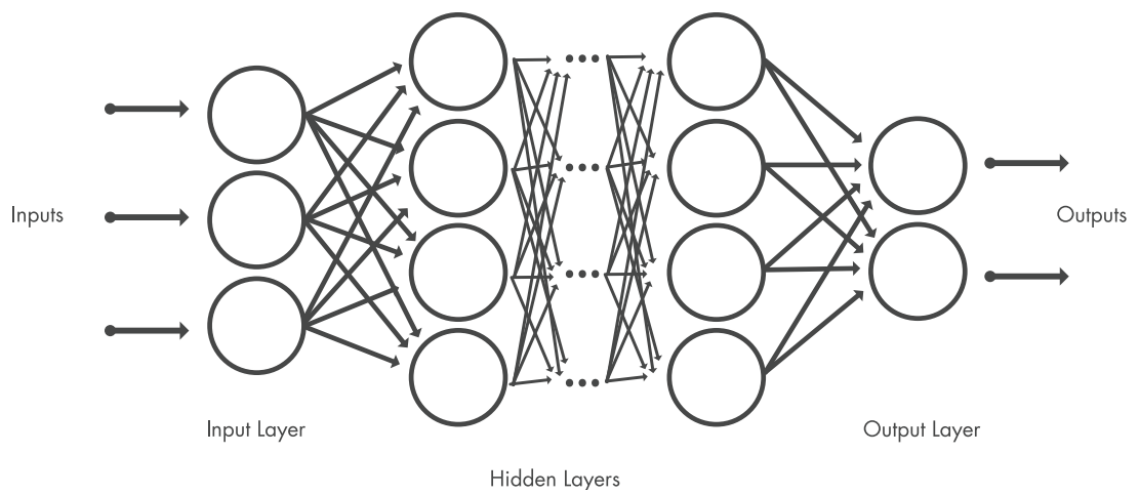
Neuronové sítě byly úspěšně použity v mnoha případech rozpoznávání vzorů, mezi které se také řadí i optické rozpoznávání osob na základě obličeje. Silnou stránkou této metody u rozpoznávání tváří je především její vysoká spolehlivost úspěšné klasifikace za předpokladu jejího správného natrénování. Úspěšnost klasifikace na bázi neuronových sítí v některých případech dokonce přesahuje úspěšnost rozpoznávání obličejů člověkem. V poslední době se často využívá tzv. hlubokých neuronových sítí založených na metodě CNN (Convolutional Neural Networks). Tato metoda bude podrobněji popsána v praktické části práce v kapitole 5.4.

#### 4.3.1 Hluboké učení – deep learning

Hluboké učení neboli deep learning je jednou z metod strojového učení, která extrahuje důležité informace (příznaky) ze vstupních dat za pomoci mnohovrstvých nelineárních výpočetních modelů,

skládajících se z jednoduchých prvků pracujících paralelně. Model se učí provádět klasifikační úkoly přímo z obrázků, textu nebo zvuku, a to především pomocí architektury neuronové sítě. Hluboké učení představuje celkový počet vrstev v síti a oproti tradičním neuronovým sítím, které mají často pouze 2 nebo 3 vrstvy, mají hluboké sítě i několik desítek skrytých vrstev.

Tyto sítě se tedy skládají ze vstupní vrstvy (vektor vstupních dat), několika skrytých vrstev a výstupní vrstvy (rozdělení objektů do jednotlivých kategorií). Vrstvy jsou vzájemně propojeny prostřednictvím uzlů nebo neuronů s každou skrytou vrstvou pomocí výstupu předchozí vrstvy jako jejího vstupu. Obecně si lze tuto architekturu představit pod následujícím obrázkem (Obrázek 32). [36] [37] [44]

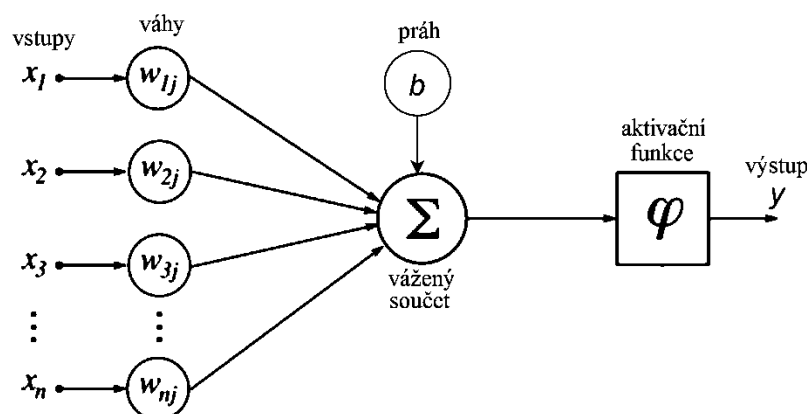


Obrázek 32: Architektura hlubokých neuronových sítí [44]

Obecně se práce s neuronovými sítěmi dělí na fázi trénovací a na fázi vybavovací. V trénovací fázi se jedná zejména o změny parametrů sítě v závislosti na trénovací sadě dat a metodě učení, např. dopředná propagace (forward pass) a propagace se zpětným šířením chyby (back propagation). V druhé fázi síť generuje výstupy klasifikováním vstupů a parametry sítě se již nemění.

#### 4.3.2 Model neuronu

Základním modelem neuronu je perceptron, jehož předloha vychází z biologického neuronu. Na vstupu do neuronu je vektor hodnot o  $n$  prvcích (Obrázek 33 vlevo), kde každá z těchto hodnot má svou váhu  $w$ , udávající její důležitost. Vážený součet těchto vstupů společně s prahem  $b$  (bias) putuje do aktivační funkce  $\phi$ . Aktivační neboli přenosová funkce je využívána k modelování nelinearity, která pomáhá neuronu vyhodnocovat komplexní problémy. Mezi nejčastěji používané aktivační funkce se řadí Heavisideův skok, Sigmoida, ReLU a hyperbolický tangens.



Obrázek 33: Model neuronu (vlastní)

### 4.3.3 Předtrénované konvoluční neuronové sítě

Trénink konvolučních neuronových sítí je velmi náročný na výpočetní výkon a to také v důsledku potřebného velkého množství trénovacích dat. V praxi se proto často využívá tzv. částečně předtrénovaných sítí (pretrained networks). Tyto sítě jsou již natrénovány na základní elementy obrazových dat a je potřeba je doučit k požadovanému účelu. Každá síť může být specifická svou architekturou, rozměrem vstupních dat atd. Mezi takovéto základní sítě patří zejména [43]:

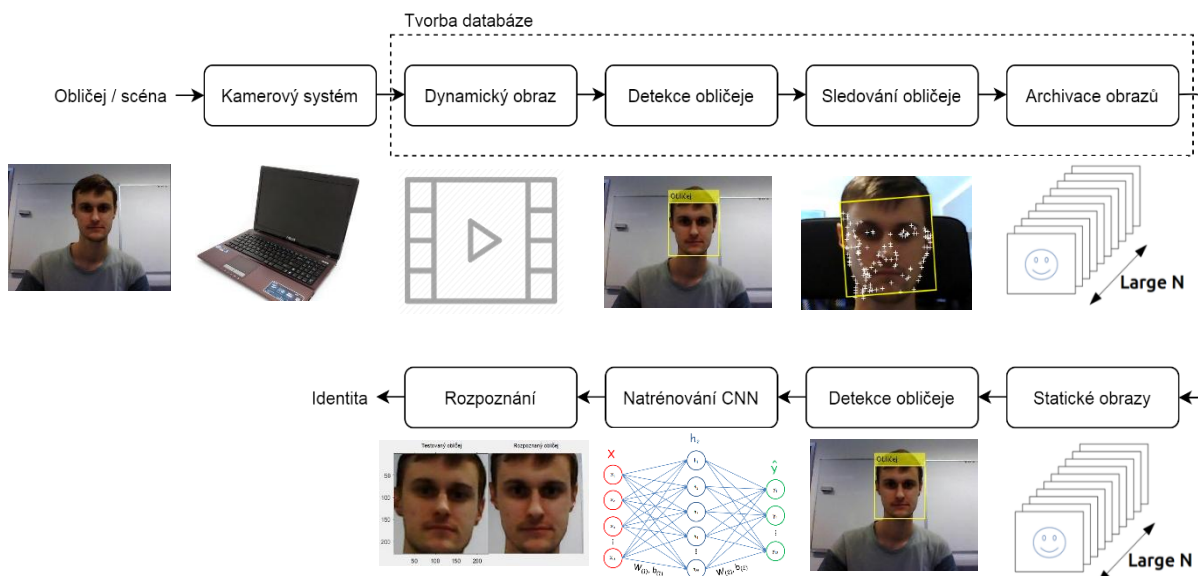
- **LeNet-5:** LeNet-5 je jednou z nejjednodušších architektur, která vznikla v roce 1998 a skládá se z 2 konvolučních a 3 plně propojených vrstev. Tato síť obsahuje okolo 60 tisíc parametrů.
- **AlexNet:** Tato síť v roce 2012 zpopularizovala metodu CNN ve strojovém vidění. Používá podobnou architekturu jako předcházející síť, ale s větším počtem skrytých vrstev a jako první zde byla použita aktivační funkce ReLU (viz kapitola 5.4.2). Počet parametrů v architektuře sítě je okolo 60 milionů.
- **VGGNet:** V roce 2014 byly společností Vision Geometry Group vyvinuty dvě sítě s různými počty vrstev, konkrétně VGG-16 a VGG-19. Tyto CNN využívaly také ReLU vrstvu stejně jako síť AlexNet, ale lišily se především velikostí konvolučních masek a pooling masek o velikostech (3x3 a 2x2). Hloubka těchto sítí se odrazila i na výsledném počtu parametrů, které převýšily dvojnásobek sítě AlexNet na přibližně 138 milionů o velikosti 500 MB.
- **GoogLeNet:** Síť GoogLeNet (2014) razantně snížila počet parametrů sítě na 4 miliony oproti předchozím dvěma sítím a svou chybovostí se přiblížila ke člověku.
- **ResNet-15:** V této síti vytvořené v roce 2015 se opět zvyšovaly počty skrytých vrstev a často docházelo k přeučení sítě. V tomto důsledku byly aplikovány metody k přeskokování spojení mezi určitými neurony. Tato síť byla také jednou z prvních, která využívala normalizaci datových sad (batch size).



## 5 Realizace a implementace algoritmů pro detekci a rozpoznávání obličejů

V této části práce bude popsán postup realizace biometrického systému, který slouží jako experimentální testovací platforma pro detekci, sledování a evaluaci metod pro detekci a rozpoznání obličejů. Bude zde popsána implementace a funkce využitých algoritmů pro detekci, sledování a rozpoznávání obličejů z obrazových dat ve vývojovém prostředí MATLAB. Obrazová data budou snímána vestavěnou webkamerou počítače a jejich zpracování bude sloužit k autentizaci osob a testování robustnosti algoritmů na různé vlivy, jako je přítomnost šumu v obraze, změna rozlišení, změna expozice nebo světelné podmínky.

Na následujícím obrázku (Obrázek 34) je znázorněn obecný postup realizace biometrického systému, kdy v prvním kroku je scéna zaznamenána webkamerou notebooku. V případě výskytu obličeje v obraze, je tento snímek uložen do databáze pro daný subjekt. Jakmile je databáze naplněna, v pořízených snímcích dané scény jsou detekovány a oříznuty obličeje. Poté je u těchto nově vytvořených snímků normalizováno rozlišení sloužící pro správné natrénování konvoluční neuronové sítě (CNN). V této chvíli je možné rozpoznávat identity daných subjektů a testovat robustnost algoritmu v závislosti na nových vstupních datech.



Obrázek 34: Postup realizace biometrického systému

### 5.1 Požadavky na snímanou scénu

Práce s obrazovými daty klade požadavky na snímanou scénu. Obecně se obraz snímáný kamerou rozděluje na oblast zájmu, tedy obličej, a pozadí. Správná detekce a rozpoznání obličeje je dosti ovlivněna velikostí obličeje v obraze, jeho pozicí, natočením, osvětlením nebo také pozadím. V mnoha případech, kdy je detekce tváře založena na barvě kůže, může pozadí s podobnou barvou značně ovlivnit kvalitu detekce.

#### Velikost obličeje v obraze

Velikostí obličeje ve snímku nebo videosekvenci se rozumí jeho rozlišení. S klesajícím rozlišením klesá také spolehlivost detekce. Proto je zapotřebí, aby měl zkoumaný obličej rozlišení minimálně 100x100 pixelů. Naopak s rostoucím rozlišením se přesnost detekce zvyšuje, avšak hlavním omezujícím

kritériem je zvyšující se výpočetní náročnost systému a je potřeba mezi těmito parametry volit kompromis.

### **Osvětlení scény**

Další důležitou roli v přesnosti rozpoznání obličejů hraje osvětlení scény. Tvář by měla být nasvícena z takového směru, aby na ní nevznikaly stíny a takovou intenzitou, aby nebyla přexponovaná a zachovávala si přirozené barvy, resp. jasovou intenzitu. Právým opakem je podexponovaný obličej, který ztrácí informace v důsledku nedostatku světla.

### **Pozice a natočení obličeje**

Aby byly algoritmy schopny detekovat a rozpoznat obličej, musí se v obraze zřetelně nacházet oči a ústa. Některé z algoritmů jsou méně citlivé na naklonění hlavy do strany a některé zase striktně vyžadují jejich přesné narovnání a zarovnání do stanovené pozice. Natáčením obličeje vůči vertikální ose dochází ke změně perspektivy, a tedy i ke změně vzdáleností mezi obličejovými rysy, což může vést ke zhoršení spolehlivosti u identifikace tváře.

### **Přítomnost šumu v obraze**

Pořizování obrazových dat není dokonalé, a to je dáno především volbou snímáče, resp. jeho kvalitou, velikostí pixelů, rozlišením, světelností, citlivostí, rychlostí závěrky a dalšími aspekty. Tyto nedokonalosti se projevují přítomností šumu v obraze, který ovlivňuje kvalitu detekce a rozpoznání tváří.

## **5.2 Komunikace webkamery a MATLABu**

Prvním krokem v implementaci algoritmů, pro rozpoznání a detekci obličejů, je zapotřebí získat obrazová data. Na internetu existuje velké množství databází, které jsou volně dostupné ke stažení. Mezi jedny z nejčastěji využívaných databází pro testování se řadí například databáze *FERET*, *Caltech Faces 1999*, *MIT-CBCL* nebo *Essex faces94* a další. Všechny tyto databáze mají různé počty subjektů, různé rozlišení obrazových dat a také odlišné pozadí. Veškeré tyto aspekty hrají při detekci a rozpoznávání tváří důležitou roli.

Druhým přístupem je vytvoření vlastních obrazových dat za pomoci webkamery nebo externí kamery, která může být k počítači připojena prostřednictvím USB (Universal Serial Bus) portu. Pro vytvoření komunikace mezi MATLABem a webkamerou je možné využít příslušných funkcí ze dvou volně dostupných hardwarových balíčků z MathWorks (Hardware Support Packages), a to *OS Generic Video Interface* nebo *USB Webcams*. Oba tyto balíčky obsahují základní funkce pro vytvoření obrazových dat ze snímáče kamery. Lze tedy jednoduše vytvořit náhled snímaných dat v podobě videa, ze kterého je pak možno snímat jak statické, tak i dynamické obrazy a ukládat je do paměti počítače pro další zpracování. V závislosti na výkonu počítače lze také takto generovaná obrazová data analyzovat v reálném čase. Ve vztahu ke komunikaci s MATLABem jsou v následující tabulce (Tabulka 2) uvedeny parametry použité webkamery v tomto projektu.

*Tabulka 2: Parametry použité webkamery*

DeviceName: 'ASUS USB2.0 WebCam'

DeviceID: 1



AdaptorName: 'winvideo'  
NativeDataType: 'uint8'  
ColorSpace 'truecolor'  
MaxWidth: 640  
MaxHeight 480

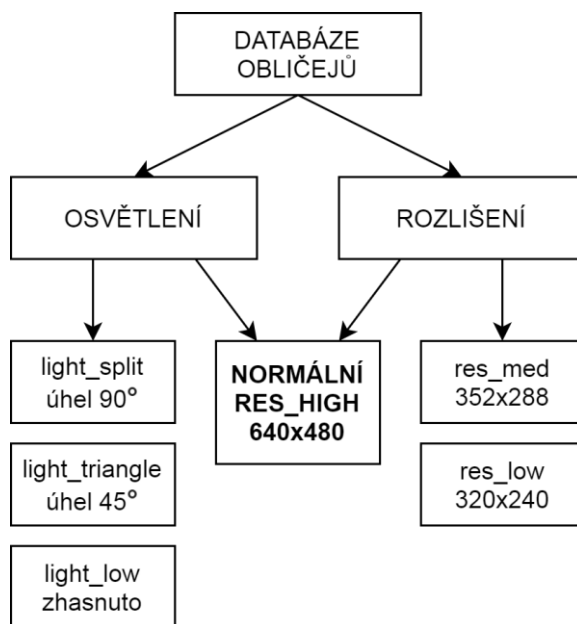
**Supported Formats:**

'YUY2\_640x480'  
'YUY2\_352x288'  
'YUY2\_320x240'  
'YUY2\_176x144'  
'YUY2\_160x120'

Různé webkamery mohou mít tyto parametry jiné a mohou umožňovat nastavení více funkcí, jako například manuální nastavení expozice nebo přesný počet snímků za sekundu u streamovaného videa.

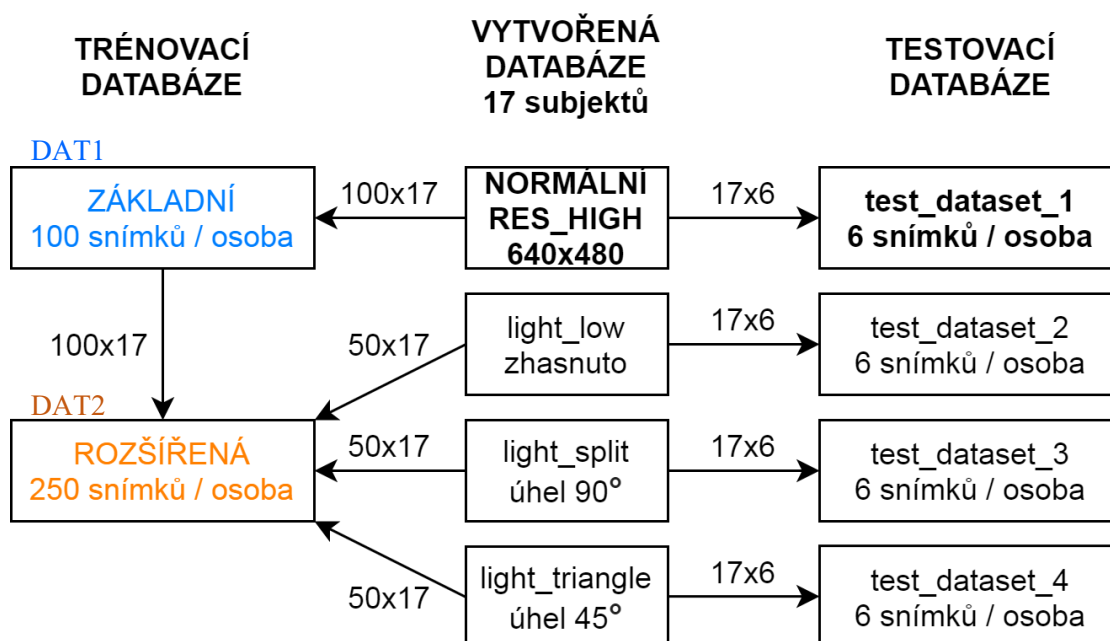
### 5.3 Tvorba databáze, detekce a sledování obličejů ve videu

V této práci byla vytvořena vlastní databáze 17 obličejů za různých podmínek. Na každou osobu připadalo 100 snímků v rámci jednoho snímání. První hlavní databáze byla snímána za nejvyššího rozlišení tj. 640x480 pixelů a za normálního ambientního osvětlení v místnosti (Obrázek 35 uprostřed). U tohoto jediného datasetu bylo snímáno 120 snímků na osobu z důvodu potřeby dalších dat pro testovací databázi. Celkem byly provedeny 4 varianty snímání při změně osvětlení a 3 varianty při změně rozlišení, datasety jsou znázorněny na Obrázek 35.



Obrázek 35: Jednotlivé datasety vytvořené databáze obličejů

Po získání potřebných dat bylo nutné tuto databázi rozdělit na data pro trénink konvoluční neuronové sítě a na obrazy sloužící pro testování systému. V následujícím obrázku (Obrázek 36) v prostředním sloupci jsou znázorněna data, ze kterých byly tvořeny dílčí databáze. Pro trénink sítě byly vytvořeny 2 datasety (Obrázek 36, vlevo). Základní databáze obsahovala veškeré snímky pořízené při normálním osvětlení a za nejvyššího rozlišení, tj. 100 snímků na osobu (DAT1). Rozšířená databáze obsahovala rovněž celou základní databázi doplněnou o 50 snímků z každého dalšího datasetu s jiným osvětlením scény (DAT2). Dále pak byly vytvořeny 4 testovací datasety. Hlavní testovací databází byl „test\_dataset\_1“, který se skládá celkem ze 102 obrazů, kde na každou osobu připadá 6 snímků. Na tento dataset byly aplikovány veškeré degradační metody uvedené v této práci (viz kapitola 6). Zbývající 3 datasety byly vytvořeny z různých osvětlení taktéž s celkovým počtem 102 snímků a byla testována robustnost detekce a klasifikace obličeje mezi těmito kategoriemi.



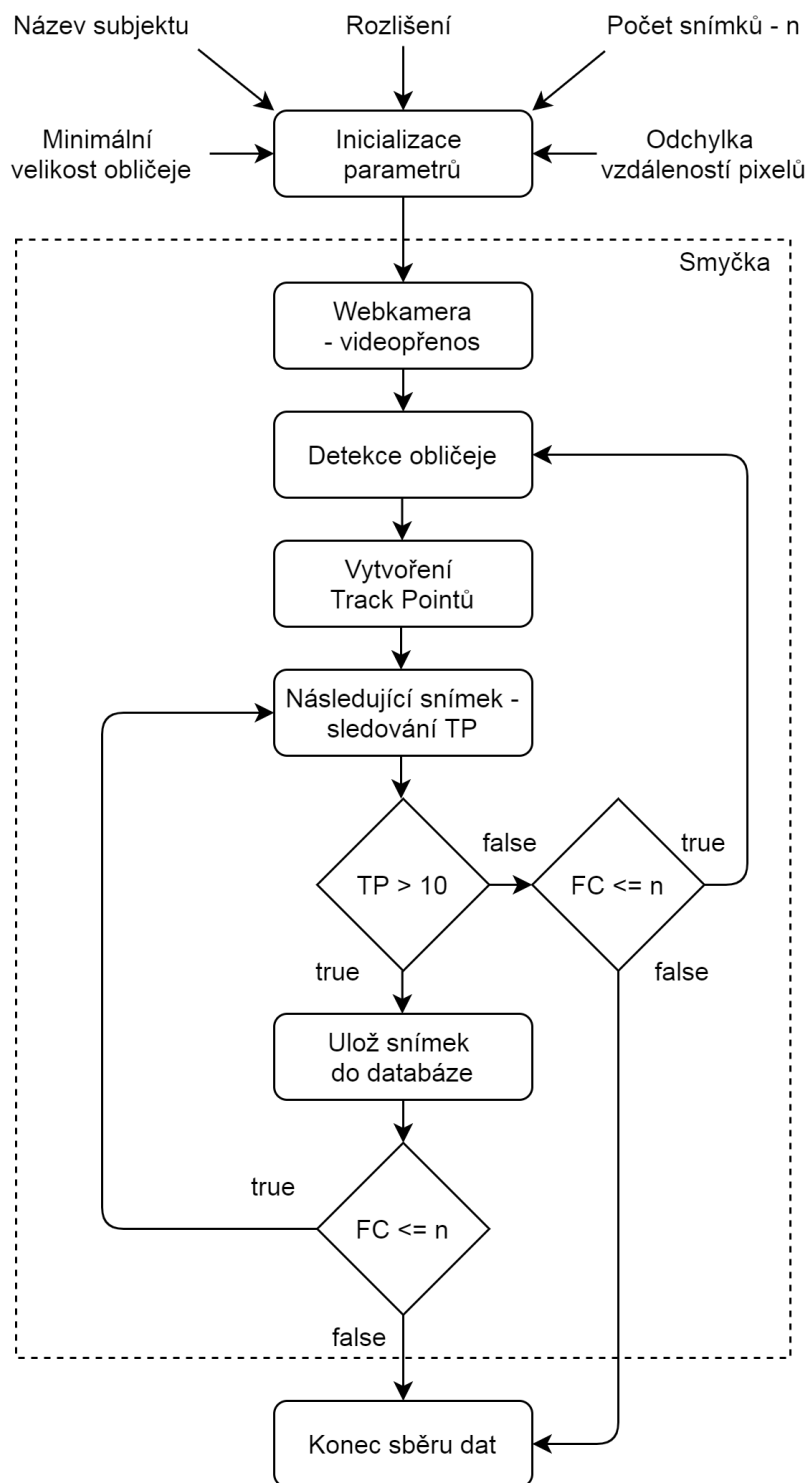
Obrázek 36: Rozdělení pořízené databáze na trénovací a testovací databázi

### 5.3.1 Algoritmus pro tvorbu databáze

Na počátku tohoto algoritmu je potřeba zadat počet snímků, ze kterých bude tvořena trénovací nebo testovací databáze, tzn. že tento počet také ovlivňuje dobu trvání algoritmu. Dále je potřeba inicializovat důležité parametry pro zobrazení a detekci obličeje v obraze. Mezi ty nejzákladnější patří rozlišení, které je detekováno automaticky za pomoci vytvoření prvního snímku (neboli snapshotu) a zjištění jeho velikosti ještě dříve, než je spuštěno video. Dalším parametrem je minimální velikost hledaného obličeje, v této realizaci se jedná o velikost 100x100 pixelů (funkce *vision.CascadeObjectDetector*) a parametr sledovací funkce, tj. maximální obousměrná odchylka o vzdálenosti 2 pixelů (funkce *vision.PointTracker*). Dále je také potřeba nastavit cestu ke složce a jméno subjektu 's1, s2, ...', pro který budou data vytvářena.

Po inicializaci parametrů a spuštění webkamery je promítán živý náhled snímané scény v barevném zobrazení 'true color' neboli RGB. V tuto chvíli se program dostává do smyčky, jejíž doba trvání je založena na již nastaveném počtu snímků pro tvorbu dat daného subjektu (Obrázek 37). V tuto chvíli zatím neexistují žádné Track Pointy (sledovací body - TP) a algoritmus musí detekovat obličej z obrazu.

Je-li obličej ve snímku detekován, jsou v této oblasti vytvořeny sledovací body na různých místech obličeje. Videosekvence postoupí o další snímek a zjišťuje přítomnost track pointů, resp. jejich počet. Pokud je tedy TP více než 10, je to známkou toho, že detekovaný obličej se v obraze stále nachází a algoritmus uloží tento snímek do databáze. Poté je čítač snímků zvýšen o jedničku (Frame Counter - FC). Pokud je FC menší nebo roven zvolenému počtu snímků  $n$ , algoritmus pokročí k dalšímu snímku sekvence a proces se opakuje.



Obrázek 37: Realizace tvorby databáze (vlastní)

V případě, že sledovací body jsou ztraceny v důsledku rychlého pohybu tváře v obraze (pokles sledovacích bodů pod 10, tj.  $TP < 10$ ), pak dojde k opětovné detekci obličeje a vytvoření nových sledovacích bodů (opět za předpokladu, že FC je menší nebo roven počtu zkoumaných snímků  $n$ ). V případě, že čítač snímků nabyde hodnoty vyšší než  $n$ , je sběr dat pro databázi ukončen. Výhodou tohoto algoritmu je, že ukládá pouze snímky, ve kterých se obličej nachází a není potřeba každý snímek sekvence podrobit detekci, nýbrž pouze sledovat body v dané oblasti obrazu.

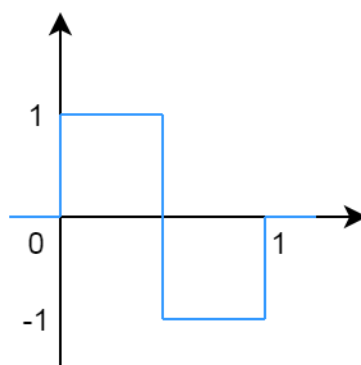
### 5.3.2 Detekce obličeje metodou Viola-Jones

Detekce metodou Viola-Jones se řadí mezi metody založené na základě vzhledu (viz kapitola 3.4). Základní myšlenka této metody se skládá ze tří částí. Využívá se zde popisu obrazu Haarovými příznaky, vytvořením integrálního obrazu a následné zapojení jednotlivých klasifikátorů do kaskády pomocí metody AdaBoost (Adaptive Boosting). Pro urychlení výpočtů této metody se často pracuje s monochromatickým obrazem, který je pro dané použití dostačující.

#### Haarovy vlnky a příznaky

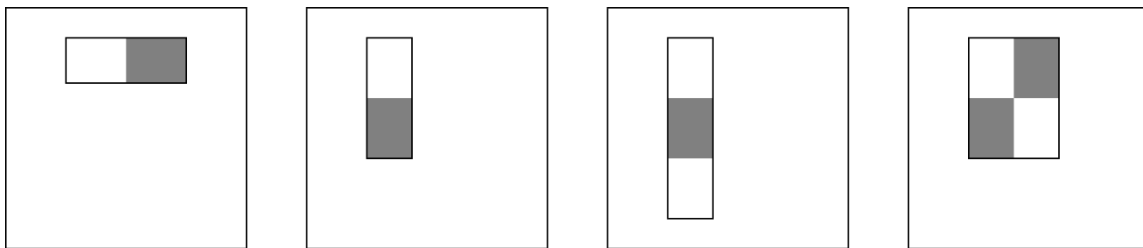
Chceme-li rychlým způsobem klasifikovat obrazové informace, je nutné charakterizovat jednotlivé oblasti v obraze co nejjednodušeji. K tomuto účelu se využívají tzv. Haarovy příznaky, které jsou založeny na stejnojmenných vlnkách (Obrázek 38), daných následujícím předpisem 5.1 [1] [21].

$$\psi(t) = \begin{cases} 1, & 0 \leq t < \frac{1}{2} \\ -1, & \frac{1}{2} \leq t < 1 \\ 0, & \text{jinde} \end{cases} \quad (5.1)$$



Obrázek 38: Haarova vlnka (vlastní)

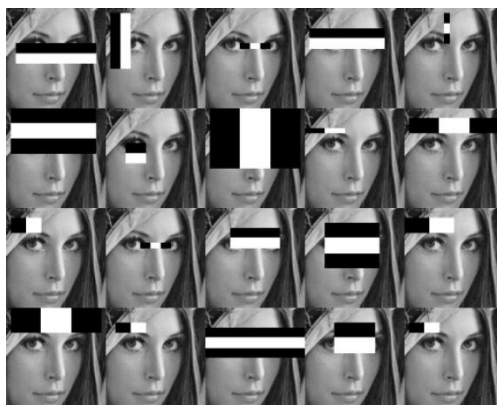
Haarovým příznakem (Obrázek 39) se rozumí hodnota intenzity rozdílu světlé a tmavé oblasti, reprezentovaných součtem jasových hodnot v daném okolí. Tento rozdíl se standardně počítá tak, že od světlých oblastí se odečítá oblast tmavá, viz vzorec 5.2, kde  $w$  (*white*) představuje hodnoty intenzit pixelů, nacházejících se ve světlé oblasti a  $b$  (*black*) hodnoty intenzit pixelů v tmavé oblasti.



Obrázek 39: Základní typy Haarových příznaků ve čtvercových oknech (vlastní)

$$f(x) = \sum_{w \in W} x(w) - \sum_{b \in B} x(b) \quad (5.2)$$

Důležitou součástí detekce Haarovými příznaky je správné nastavení jejich parametrů. Mezi tyto parametry patří zejména rozsah prověřované oblasti, délka posunu v jednotlivých krocích a minimální / maximální velikost oblasti (jak již bylo zmíněno výše, v realizovaném řešení se jedná především o definování minimální velikosti oblasti obličeje na 100x100 pixelů). Všechny tyto parametry musí být nastaveny v závislosti na velikosti vstupního obrazu tak, aby byl obraz prověřen co nejdetailněji, avšak za co nejkratší čas. Příklad použití Haarových příznaků je znázorněn na obrázku (Obrázek 40).



Obrázek 40: Reálné použití Haarových příznaků v obraze [21]

Pokud jsou Haarovy příznaky posouvány po obraze s příliš velkým krokem, časová náročnost klesá, avšak některé oblasti v obraze nemusí být správně prověřeny. Na druhou stranu, je-li posun Haarových příznaků velmi malý, dostáváme záruku precizně prověřeného snímku, ale samotná operace bude trvat příliš dlouhou dobu.

Dalším důležitým parametrem je také koeficient zvětšení, na jehož základě mění samotný příznak svou velikost, která slouží k přesnější detekci v případech, že daná osoba stojí ve větší nebo menší vzdálenosti od snímáče. V tomto řešení byl koeficient zvětšení nastaven na hodnotu  $ScaleFactor = 1,1$ , což znamená, že v každém kroku zvětšení se oblast Haarových příznaků zvětšuje o 10 %. Velikost tohoto koeficientu musí být větší nebo roven hodnotě  $ScaleFactor = 1,0001$ . Nesprávným nastavením koeficientu zvětšení společně s velikostí kroku může dojít k vícenásobnému detekování oblasti obličeje v obraze.

Tento nežádoucí jev lze potlačit nastavením slučovacího prahu (*MergeTreshhold*), který je dán celočíselnou hodnotou. Tato vlastnost definuje minimální počet detekcí v různých škálách a pozicích k deklarování finální detekce obličeje. Jinými slovy, jednotlivé detekce obličejových rysů se sloučí a vytvoří jedno pole kolem celého cílového objektu. Zvýšením této prahové hodnoty je tedy potlačen vícenásobný výskyt obličeje ve stejné oblasti obrazu. V algoritmu byl tento parametr nastaven na hodnotu *MergeThreshold* = 4.

## Integrální obraz

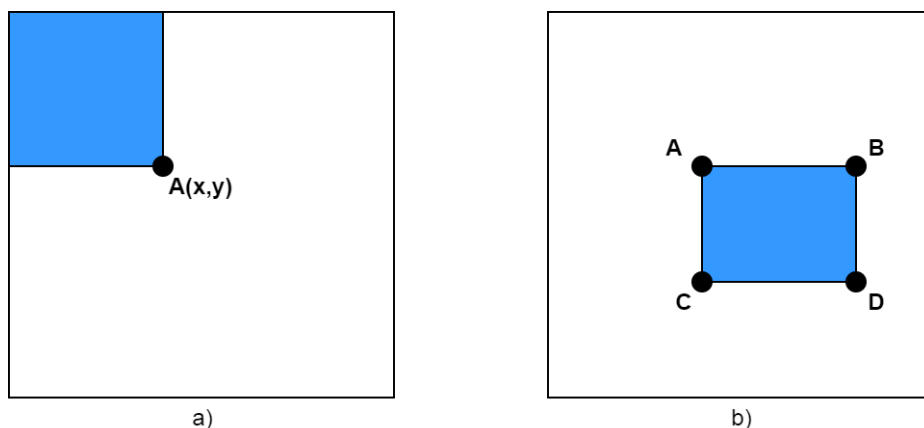
Jak již bylo zmíněno, časová náročnost je také často jedním z hlavních faktorů u zpracování obrazu. Pro snížení výpočetního času se vstupní obraz převádí na tzv. obraz integrální, který dokáže efektivně spočítat hodnoty jednotlivých Haarových příznaků. Tyto hodnoty nejsou počítány z celého nově vytvořeného integrálního obrazu, nýbrž pouze ze čtvercové okenní oblasti, která se po obraze posouvá.

Hodnoty integrálního obrazu jsou dány vztahem 5.3, kde  $ii(x,y)$  je hodnota intenzity pixelu integrálního obrazu a  $i(i,j)$  je hodnota intenzity pixelu původního obrazu.

$$ii(x,y) = \sum_{i,j=0}^{i,j \leq x,y} i(i,j) \quad (5.3)$$

Nová hodnota  $A(x,y)$  tedy vzniká součtem intenzit bodů, které jsou směrem nahoru a doleva od pozice počítaného pixelu, viz Obrázek 41a (tzn., že úplně poslední bod v obraze vpravo dole představuje součet všech pixelů obrazu). Samotný výpočet obdélníkové plochy dané oblasti (Obrázek 41b) v obraze již není náročný, vzhledem k tomu, že se jedná o pouhé 3 matematické operace, popsány v rovnici 5.4.

$$S = A + D - B - C \quad (5.4)$$



Obrázek 41: Integrální obraz: a) hodnota bodu A reprezentující modře vyznačenou oblast původního obrazu, b) plocha S, reprezentována modrou barvou, vycházející z rovnice 5.4 (vlastní)

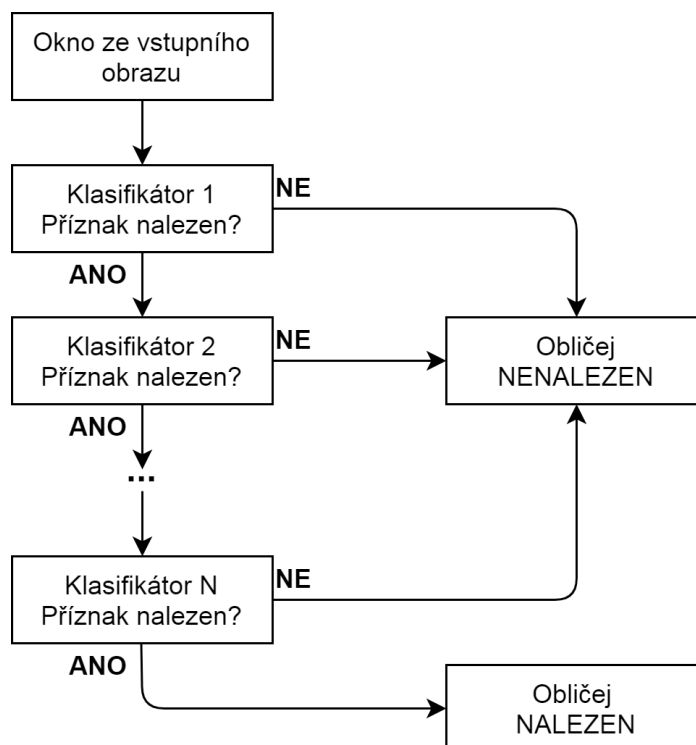
## Klasifikátor AdaBoost

AdaBoost (z ang. Adaptive Boosting) je učící se klasifikátor kombinující vhodným způsobem velké množství jednoduchých, málo spolehlivých klasifikátorů do jednoho obsáhlého robustního klasifikátoru. Hlavní výhodou tohoto klasifikátoru je, že šetří výpočetní výkon a urychluje průběh detekce. To proto, že v případě, kdy se obličej v obraze nevyskytuje, je mnohem vyšší šance zavrhnutí tohoto předpokladu na základě jednoho z prvních klasifikátorů v kaskádě. Jak už bylo zmíněno výše, snímek není analyzován celý najednou, nýbrž postupně pomocí posuvných oken.

Trénování komplexního klasifikátoru probíhá pomocí sady slabých klasifikátorů v následujících fázích:

1. Nastavení vah pro tréninková data na výchozí hodnotu – mohou být náhodně zvolená čísla
2. Výběr klasifikátoru s nejlepšími výsledky na tréninkových datech (z jednotlivých slabých klasifikátorů s ohledem na jejich váhy).
3. Výpočet koeficientu pro vybraný klasifikátor vzhledem k jeho úspěšnosti.
4. Úprava vah u trénovacích dat. Váha roste pro data, která byla špatně klasifikována, a klesá pro data, která byla klasifikována správně.
5. Přesun k bodu č. 2 dokud není komplexní výsledný klasifikátor dostatečně přesný nebo není dosažen předem daný počet iterací.

Kaskádním řazením klasifikátorů docílíme rychlé detekce daného objektu v obraze. Jednotlivé části vstupního snímku prochází postupně řadou klasifikátorů, které hledají dané příznaky a vyhodnocují jeho vlastnosti. Nejjednodušší klasifikátory jsou tedy zařazeny na začátek tohoto procesu, a to proto, aby byla detekce efektivní zejména v případě, kdy se na vstupu tvář nevyskytuje a výsledek mohl být znám okamžitě, aniž by testoval další příznaky (obličej NENALEZEN), viz Obrázek 42. Pokud dané okno snímku projde všemi klasifikátory algoritmu úspěšně, je v této oblasti detekován obličej (obličej NALEZEN).



Obrázek 42: Kaskádní řazení klasifikátorů (upraveno podle [1])

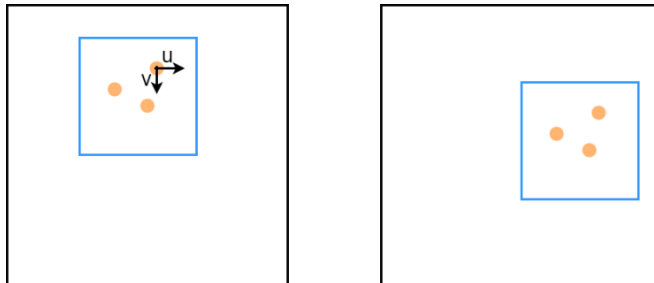
Na stejném principu se dá také detekovat mnoho dalších vlastností, např. pozice očí a úst. V případě nalezení obličeje v obraze je možné postoupit k dalšímu kroku a tím je jeho sledování a následné rozpoznání obličeje.

### 5.3.3 Sledování obličeje metodou KLT

Jakmile získáme informaci o detekovaném obličeji, je možné sledovat pohyb významných bodů tváře (tzv. příznaky jednoduché pro sledování - „*good features to track*“) pohybujících se v obraze. Mezi takovéto příznaky se řadí nejčastěji hrany v obraze. K tomuto účelu je v tomto projektu využit trackovací algoritmus KLT (Kanade-Lucas-Tomasi). Tento algoritmus může být aplikován na aktuální snímky bez nutnosti nějakého předtrénování, záleží pouze na informaci o sledovaných bodech v předcházejícím snímku přehrávaného videa. Tento algoritmus v navrženém systému pracuje s monochromatickým obrazem pro ušetření výpočetního výkonu, avšak vyobrazení detekovaných bodů je zobrazeno v barevném obraze.

Tato metoda spolehlivě funguje na základě následujících tří předpokladů:

- **Malý pohyb.** Body extrahovaných rysů mají od sebe krátkou vzdálenost mezi dvěma sousedními snímky. Posun bodů může být popsán 2D vektorem  $[u, v]^T$ , znázorněno na následujícím obrázku (Obrázek 43).



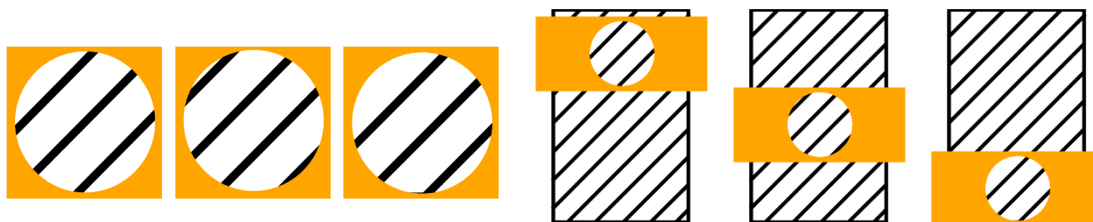
Obrázek 43: Posun extrahovaných bodů metody optical flow (vlastní podle [32])

Jedná se o tzv. maximální obousměrnou odchylku (z ang. Maximum Bidirectional Error). Tato odchylka udává vzdálenost v pixelech mezi jednotlivými příznaky ve dvou sousedních snímcích videa. V navrženém systému byla tato hodnota nastavena na 2 pixely. V případě, že tato hodnota je mezi snímky vyšší než uvedená prahová hodnota, pak jsou sledovací body ztraceny. Použití obousměrné chyby je tedy efektivním způsobem, jak eliminovat příznaky, které nelze spolehlivě sledovat.

- **Stálost jasu.** Projekce stejných bodů v obraze musí mít téměř stejný jas v každém snímku videa. To znamená, že u monochromatických obrazů se předpokládá, že se jasové hodnoty pohybujících se pixelů daného objektu nebudou měnit v závislosti na čase (uvažují se konstantní uměle nastavené světelné podmínky pro danou scénu).



- **Prostorová koherence.** Tato vlastnost předpokládá, že sousední body, které patří ke stejnému objektu v obraze, mají podobný pohyb a promítají se do blízkých oblastí v obrazové rovině (viz oranžové body, Obrázek 43). Přesto se zde může vyskytovat tzv. aliasing, který způsobuje špatný odhad pohybu jednotlivých oblastí. Tento jev je znázorněn na Obrázek 44, kdy na levém snímku není možné z jednotlivých vzorků identifikovat směr pohybu čar v obraze. Kdežto na obrázku vpravo je znázorněn skutečný pohyb čar v obraze, které se v sekvenci vzorků pohybují směrem nahoru.



Obrázek 44: Aliasing: vlevo neznámý pohyb, vpravo skutečný pohyb [33]

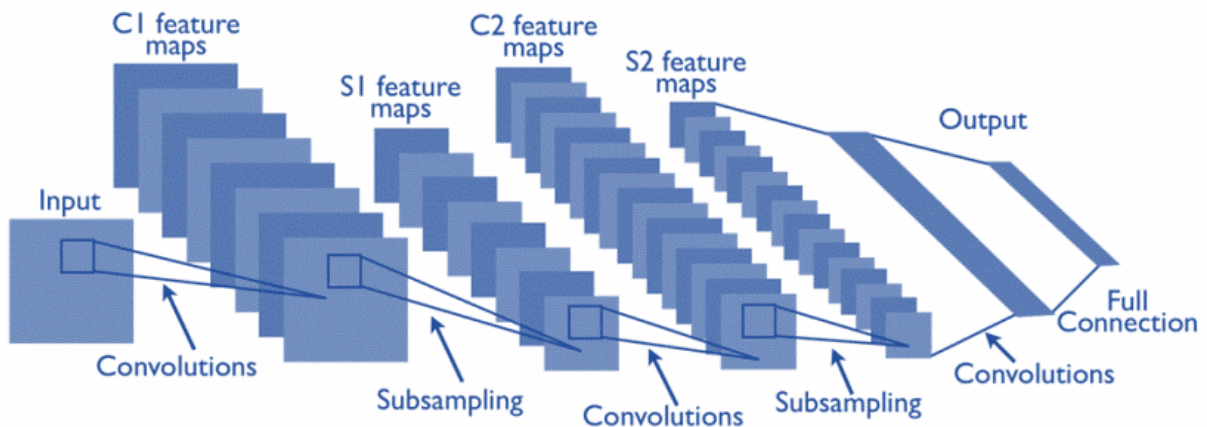
## 5.4 Realizace systému pro rozpoznání identity člověka na základě obličeje

Pro rozpoznávání obličejů byla aplikována metoda konvolučních neuronových sítí s předtrénovanou sítí AlexNet. Tato metoda se řadí do tzv. hlubokého učení, které si lze představit pod architekturou, která je složená ze vstupní vrstvy, několika skrytých vrstev a výstupní vrstvy (viz kapitola 4.3.1, Obrázek 32). Tato síť byla dotrénována ze statických snímků a také aplikována na datasey se statickými obrazy pro účely klasifikace osob podle obličejů.

### 5.4.1 Konvoluční neuronová síť – CNN

Konvoluční neuronové sítě jsou jedním z odvětví neuronových sítí a disponují mnohem menším počtem parametrů než klasické neuronové sítě. Tento fakt je dán především využitím principu konvoluce a sdružování neboli podvzorkování (z ang. pooling). Díky tomu je možné zpracovávat data o velkých rozměrech, a proto se tato metoda uplatňuje především u zpracování obrazových dat. Konvoluční neuronové sítě se skládají z několika vrstev, kterými vstupní data resp. jejich příznaky procházejí.

Mezi hlavní funkce této architektury (zobrazené na Obrázek 45) patří vrstva konvoluční (convolution), aktivační (ReLU) a sdružovací (pooling nebo také subsampling). Tyto 3 vrstvy se několikrát opakují a rozkládají obraz na příznaky, které analyzují. Výsledek tohoto procesu se promítne do předposlední tzv. plně propojené vrstvy, která vytváří spojení mezi každým vstupním prvkem a každým výstupním prvkem vektoru. Klasifikační vrstva řídí finální rozhodnutí o klasifikaci dané osoby na základě míry důvěry sítě ve správnost klasifikace.



Obrázek 45: Architektura konvoluční neuronové sítě [35]

#### 5.4.2 AlexNet

AlexNet je předtrénovaná konvoluční neuronová síť, která dokáže rozeznávat různé předměty a je natrénována ze 1,2 milionů obrazů. Jedná se o architekturu supervizorního učení, neboli učení s učitelem, které vyžaduje apriorní informace. To znamená, že vstupní data musí mít předem stanovené správné zařazení do klasifikační třídy (viz kapitola 3.4). Síť AlexNet se skládá z 23 vrstev (resp. 25 vrstev včetně klasifikačních vrstev) s normalizovaným vstupem o rozměrech 227x227 pixelů pro 3 chromatické složky barevného modelu RGB.

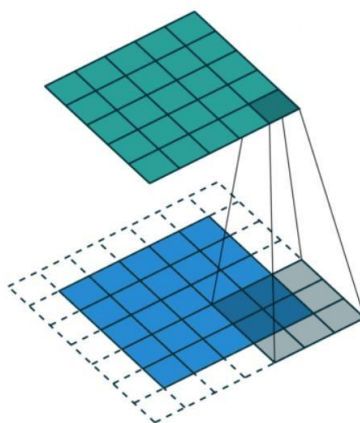
#### 5.4.3 Vrstvy sítě AlexNet

Jak již bylo zmíněno, vstupní data prochází jednotlivými vrstvami sítě. V této podkapitole budou popsány funkce hlavních vrstev sítě. Kompletní seznam všech 25 vrstev sítě AlexNet je možné najít v odkaze [38].

- **Konvoluční vrstva**

V této vrstvě je na vstup nebo na předešlou vrstvu sítě aplikována operace konvoluce. Konvoluce přikládá ke vstupnímu obrazu konvoluční masku (jádro) o daných rozměrech např. u první konvoluční vrstvy (conv1) je prováděno 96 konvolucí s velikostí jádra 11x11x3, kde první dvě čísla představují šířku a výšku masky a třetí číslo hloubky jádra (počet chromatických vrstev). Jednotlivé buňky jádra ležící na sobě se násobí se vstupním obrazem, provede se jejich součet a výsledná hodnota se ukládá do nového obrazu. V dalším kroku se maska posouvá o předem stanovenou vzdálenost (šířka kroku neboli stride) v pixelech a výpočet se opakuje. V první konvoluční vrstvě je tento krok nastaven na hodnotu [4 4], což znamená, že se konvoluční maska posouvá v horizontálním směru vždy o 4 pixely a ve vertikálním směru taktéž o 4 pixely v jednotlivých krocích konvoluce. Dalším parametrem u konvoluce je tzv. padding, který je v první vrstvě nastaven na [0 0 0 0]. Padding neboli šířka rámu vstupní matice, udává počet řádků nebo sloupců, o které je vstupní matice rozšířena a vyplněna nulami. Vzhledem k situaci, že v první vrstvě je rozšíření nulové, dojde vlivem konvoluce ke zmenšení obrazu.

Obrázek 46 znázorňuje princip konvoluce, kde modrá barva představuje vstupní snímek a zelená barva výstupní matici. Konvoluční jádro je v tomto případě  $3 \times 3 \times 1$  se střídou 1. Na obrázku je také znázorněný padding, kde hrany původního obrazu jsou rozšířeny o pixely s hodnotou 0 (prázdné buňky okolo původního obrazu). Nová matice oproti původní tedy není zmenšena.



*Obrázek 46: Znázornění konvoluce, jejího jádra a paddingu*

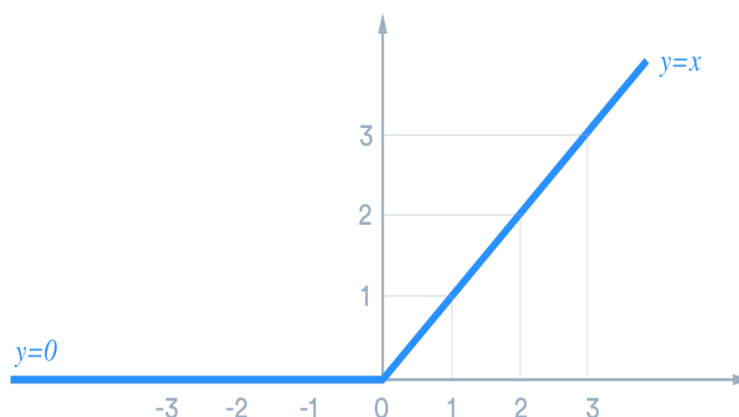
V architektuře sítě AlexNet je celkem 5 konvolučních vrstev s různým počtem konvolucí, s různou velikostí konvolučního jádra, posunu a paddingu. Kompletní přehled lze najít na již zmíněném odkaze [38].

- **Aktivační vrstva ReLU**

Po konvoluční vrstvě následuje vrstva aktivační, konkrétně ReLU vrstva (Rectified Linear Unit layer). Aktivačních nebo také přenosových funkcí neuronu je několik a mezi ty nejčastější patří sigmoid, hyperbolický tangens a také ReLU. Aktivační funkce se používá k určení výstupu neuronu. V závislosti na zvolené funkci mapuje výsledné hodnoty mezi 0 až 1 (sigmoid) nebo -1 až 1 (tanh). Přenosovou funkcí může být také Heavisideova skoková funkce.

Při práci s CNN se nejvíce osvědčila právě funkce ReLU, která je v mnoha podobných řešeních využívána a její funkce je definována vztahem 5.5 a znázorněna grafem (Obrázek 47).

$$y(x) = \max(0, x) \quad (5.5)$$



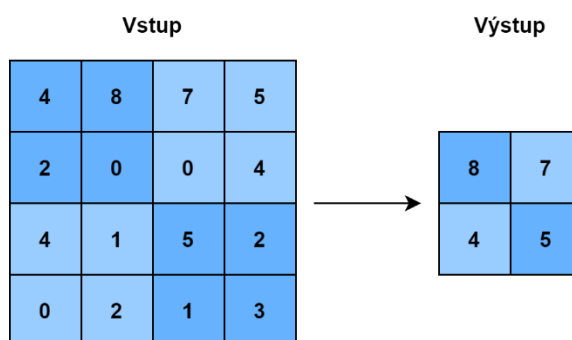
Obrázek 47: Přenosová funkce ReLU [39]

Mezi hlavní výhody ReLU oproti ostatním funkcím v CNN patří především její výpočetní efektivita a nedochází u ní k problémům se saturací.

- **Sdružovací funkce – Pooling**

Další vrstvou v síti je tzv. sdružovací funkce, která má za úkol podvzorkovat a tím zredukovat výstupní data z předcházejících vrstev. Výsledkem tohoto procesu je zlepšení distribuce parametrů a snížení výpočetního času v průběhu trénování sítě. U sdružovací funkce se podobně jako u konvoluce volí maska o daném rozměru, krok (jak horizontální tak vertikální) a padding. V síti AlexNet se konkrétně využívá funkce MaxPooling, která vybere maximální hodnotu z matice pod daným jádrem a zapíše ji do výstupní matice.

Ukázka sdružování MaxPooling s jádrem 2x2, střídou [2 2] a bez paddingu je zobrazena na obrázku níže (Obrázek 48). V tomto případě dochází k redukci dat o 75 %.



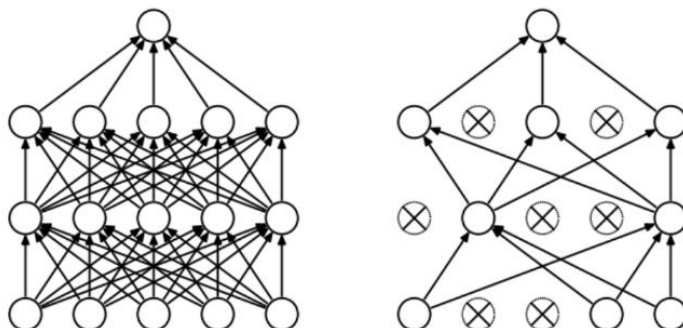
Obrázek 48: Sdružování funkcí Max Pooling

Ve všech případech použití sdružování v této síti je maska velká 3x3 pixely s krokem [2 2] a nulovým paddingem.

- **Dropout**

Vrstva dropout má za úkol přerušit některé vazby mezi náhodnými neurony, během trénování sítě, s předem určenou pravděpodobností (Obrázek 49). To znamená, že jejich příspěvek k aktivaci neuronů je dočasně odstraněn při dopředném průchodu (forward propagation) a žádné změny vah nejsou aplikovány na neuron při zpětné propagaci chyby (backward propagation).

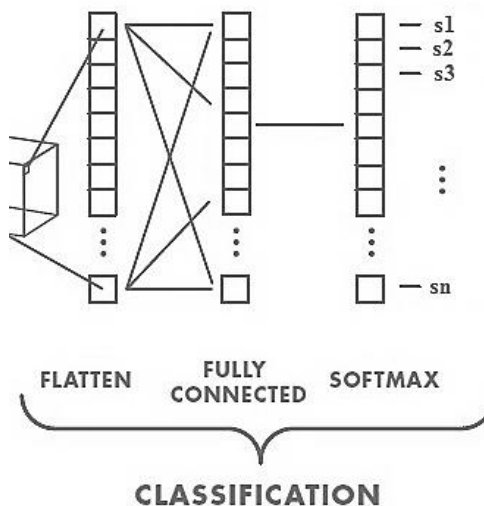
Výsledkem je tedy síť, která se stává méně citlivou na specifické váhy neuronů, a tudíž je zde menší pravděpodobnost, že by došlo k jejímu přeučení. Vliv vah je tedy rozprostřen rovnoměrněji, a to pozitivně ovlivňuje výsledek predikce.



Obrázek 49: Proces přerušení spojení neuronů ve vrstvě sítě dropout [40]

- **Plně propojená vrstva – Fully Connected layer**

Cílem plně propojené vrstvy (FC) je vzít výsledky procesu konvoluce a sdružování a použít je ke klasifikaci obrazu do tříd. Výstup konvoluční a sdružovací vrstvy je zploštěn do jediného vektoru hodnot (Obrázek 50), z nichž každá představuje pravděpodobnost, že určitý prvek patří k dané výstupní klasifikační třídě. Tato vrstva tedy propojí každý vstupní prvek s každým výstupním prvkem a tím zlepšuje distribuci požadované informace.



Obrázek 50: Klasifikace CNN: FC vrstva, Softmax vrstva, klasifikační třídy (upraveno podle [44])

- **Softmax vrstva**

Vrstva Softmax je implementována prostřednictvím vrstvy neuronové sítě hned za FC vrstvou a těsně před výstupní vrstvou. Vrstva Softmax je aktivační funkcí a musí mít stejný počet uzlů jako výstupní vrstva. Aplikace této funkce zajišťuje, že jednotlivé buňky vektoru budou nabývat rozsahu hodnot od 0 do 1 (podm. 5.7) a jejich součet bude roven 1 (podm. 5.8). Tyto hodnoty jsou vypočítány na základě vah vektoru plně propojené vrstvy.

Jak již bylo zmíněno, jedná se o aktivační funkci stejně jako je sigmoida, hyperbolický tangens nebo ReLU a je definována vztahem (5.6):

$$y_r(x) = \frac{e^{a_r(x)}}{\sum_{j=1}^k e^{a_j(x)}} \quad (5.6)$$

Kde  $k$  je počet tříd,  $x$  je vstupní vektor a  $y_r(x)$  je pravděpodobnost s danou klasifikační třídou. Pro danou rovnici musí platit tyto podmínky (5.7) a (5.8):

$$0 \leq y_r \leq 1 \quad (5.7)$$

$$\sum_{j=1}^k y_j = 1 \quad (5.8)$$

- **Klasifikační vrstva – Classification Output layer**

Klasifikační vrstva vypočítává ztrátovou funkci „*Cross entropy loss*“ pro klasifikace se vzájemně se vylučujícími třídami. U typických klasifikačních sítí musí klasifikační vrstva následovat vrstvu softmax. Výstupem této vrstvy je klasifikační třída, do které spadá vstupní obraz, tj. jméno konkrétního subjektu (např. „s1“, „s2“, atd.).

#### 5.4.4 Učení CNN

Učení konvoluční neuronové sítě od samotného počátku může i na velmi výkonných počítačích zabrat dny až týdny. Proto bylo v tomto řešení využito předtrénované sítě AlexNet, kterou bylo zapotřebí dotrénovat k požadovanému účelu. Za účely testování míry přesnosti klasifikace byla síť doučena dvěma různými databázemi. Jednalo se o základní a rozšířenou databázi (viz kapitola 5.3, Obrázek 36).

Učení sítě tedy probíhalo na základě snímků, ve kterých byl oříznut pouze obličej, a které byly normalizovány rozlišením na 227x227x3. Trénovací databáze byla dále v programu náhodně rozdělena na trénovací část a testovací neboli kontrolní část v poměru 0,8 : 0,2. Kontrolní část databáze poskytuje síti informace o přesnosti predikovaných obličejů a slouží k optimalizaci vnitřních parametrů sítě. Hlavním úkolem je minimalizovat chybu predikce systému.

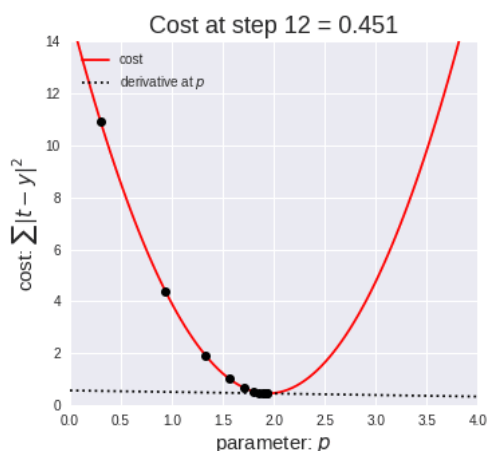
#### Nastavení pro trénink sítě:

Pro lepší pochopení těchto nastavení je nutné prvně vysvětlit, co je to tzv. Gradient Descent.

#### Gradient Descent

Gradient descent (sestupový gradient) je iterační optimalizační algoritmus používaný ve strojovém učení k nalezení nejlepších výsledků. Gradient znamená míru naklonění nebo sklonu svahu. Descent určuje směr, tedy směr sestupný. Algoritmus je iterativní, což znamená, že musíme získat výsledky vícekrát, abychom získali co nejoptimálnější výsledek. Iterativní kvalita sestupového gradientu pomáhá síti se optimálně přizpůsobit vstupním datům. Gradient descent má parametr nazvaný jako learning rate

(rychlost učení). Na obrázku (Obrázek 51) jsou jednotlivé kroky znázorněny černými tečkami a vzdálenost mezi nimi udává právě měnící se learning rate.



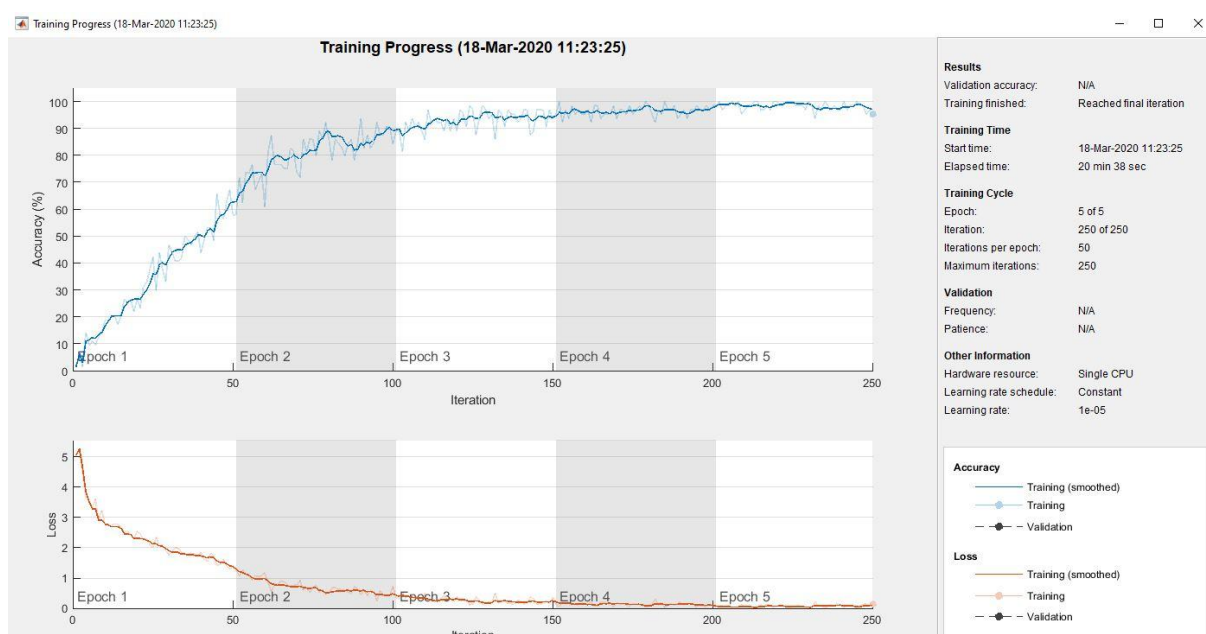
Obrázek 51: Gradient descent [42]

- **Learning rate:** Tento parametr určuje velikost změny modelu během každého kroku tohoto vyhledávacího (optimalizačního) procesu a nazývá se jako „rychlost učení“. Rychlost učení je jedním z nejdůležitějších parametrů pro vyladění neuronové sítě, aby se dosáhlo dobrého výkonu u řešeného problému. Obecně, čím je tento parametr menší, tím roste míra spolehlivosti klasifikace CNN, ale zároveň se prodlužuje doba natrénování sítě.
- **RMS Propagation:** Jedná se o typ sestupného gradientu s názvem střední kvadratická propagace (z ang. Root Mean Squared Propagation), která sleduje vážený průběžný průměr kvadratického gradientu a poté rozděluje každý vypočtený gradient druhou odmocninou tohoto váženého průběžného průměru (v podstatě normalizuje gradient dělením velikostí předchozích gradientů). Důsledkem je, že v případě kdy je gradient velmi malý, aktualizace podniknou větší kroky, a tím zajistí rychlejší učení sítě. [41]
- **Max Epochs = 5:** Maximální počet epoch při trénování. Jedna epocha je proces, při kterém celý datový soubor projde neuronovou sítí dopřednou a zpětnou propagací pouze jednou. Jelikož využíváme pouze omezený soubor dat a iterativně optimalizujeme učení za pomoci Gradient descent, tak jediná epocha pro naučení sítě nestačí (tzv. underfitting). Pro vytvoření kvalitní sítě je tedy potřeba aktualizovat váhy neuronů vícekrát a proto je potřeba aby trénovací databáze prošla tréninkem sítě ve více epochách. Jakmile se počet epoch zvětšuje, zvyšuje se i přesnost klasifikace sítě, avšak může dojít k problému přeučení při volbě nadměrného množství těchto cyklů. Obecný ideální počet epoch neexistuje, protože vše souvisí s množstvím datové sady a různorodostí dat v této sadě.

- **Mini Batch Size:** Vzhledem k situaci, že není možné předat celý datový soubor do neuronové sítě najednou, je potřeba jej rozdělit do počtu šarží, jinými slovy rozdělit jej na části.
- **Num of iterations:** Iterace je počet šarží (Batch) potřebných k dokončení jedné epochy učení.

Postup trénování sítě pro rozšířenou databázi DAT2 je znázorněn v grafu na obrázku níže (Obrázek 52). Trénování probíhalo v 5 epochách s nastavenou rychlostí učení na 0,00001. Horní graf zobrazuje míru přesnosti rozpoznání testovací části dat z trénovacího datasetu a dolní graf popisuje průběh ztrátové funkce.

Proces trénování základní databáze DAT1, se 100 obrazy na osobu, trval přibližně 11 minut a pro rozšířenou databázi o velikosti 250 snímků na osobu přibližně 21 minut. Trénování proběhlo na počítači s procesorem Intel Core i7 2630QM s taktovací frekvencí 2.0 GHz, operační paměť 8 GB a bez podpory grafické karty GPU.



Obrázek 52: Proces trénování sítě AlexNet rozšířenou trénovací databází DAT2



## 6 Testování a vyhodnocení robustnosti realizovaného biometrického systému

Při realizaci jakéhokoliv systému je potřeba objektivně otestovat jeho kvalitu. Pro tyto účely byly vytvořeny 4 testovací databáze obsahující 102 obrazů (kapitola 5.3). Hlavním testovacím datasetem pro klasifikaci obličejů byla databáze „*test\_dataset\_1*“, na kterou byly aplikovány různé degradační metody. Proběhlo také srovnání mezi kvalitou detekce obličejů pomocí algoritmu Viola-Jones a kvalitou detekce obličejů sledovacím algoritmem KLT. Veškerá tato srovnání budou uvedena a zhodnocena v této kapitole. Přesné hodnoty výsledků lze nalézt v tabulkách v příloze.

### 6.1 Míra přesnosti detekovaných obličejů sledováním a jednotlivou detekcí

V této podkapitole bude srovnána četnost detekovaných obličejů sledováním a jednotlivou detekcí ve snímcích z realtime videa při různém rozlišení a osvětlení.

V nadcházející tabulce (Tabulka 3) jsou uvedeny počty obrazů, které byly celkově snímány při tvorbě databáze obličejů v laboratoři. V levé části tabulky jsou také uvedené rozměry rozlišení a světelné podmínky, které byly při snímání aplikovány. Je nutné podotknout, že byla očekávána větší přesnost detekce sledovacím algoritmem, která se také potvrdila. To proto, že u sledovacího algoritmu nehraje významnou roli natočení hlavy tak jako u samotné detekce kaskádním klasifikátorem.

*Tabulka 3: Srovnání četnosti detekovaných obličejů sledováním a jednotlivou detekcí obličejů*

<b>ROZLIŠENÍ (light_high)</b>	Počet snímaných obrazů	Počet snímků s obličejem trackováním	Počet snímků s obličejem detekcí	Míra detekovaných obličejů trackováním (%)	Míra detekovaných obličejů detekcí (%)
640x480	2040	2014	1892	<b>98,73</b>	<b>92,75</b>
352x288	1700	1650	1440	<b>97,06</b>	<b>84,71</b>
320x240	1700	1559	1014	<b>91,71</b>	<b>59,65</b>
<b>SVĚTELNÉ PODMÍNKY</b>					
light_split	1700	1620	1537	<b>95,29</b>	<b>90,41</b>
light_triangle	1700	1585	1530	<b>93,24</b>	<b>90,00</b>
light_low	1700	1593	1493	<b>93,71</b>	<b>87,82</b>

V tabulce lze pozorovat, že při změně velikosti obrazu (při normálním osvětlení), detekce trackováním neklesá pod 90 %. To stejné se ovšem nedá říct o jednotlivé detekci, kde je pokles detekovaných obličejů razantně větší. Situace může být také ovlivněna probandy, kteří se nemuseli před kamerou pohybovat vždy úplně stejně a mohli se od snímače kamery vzdálit dále než při měřeních s vyšším rozlišením.

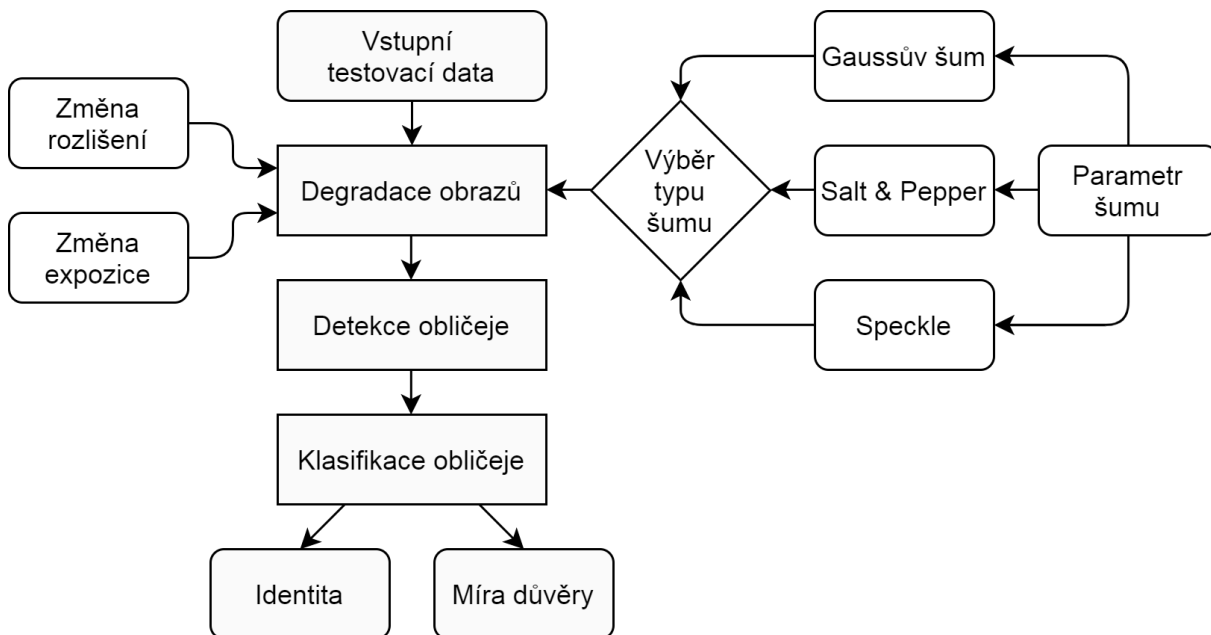
Každopádně to nemění nic na faktu, že samotná detekce je značně méně spolehlivá než sledovací algoritmus.

Při změně světelných podmínek s nejvyšším rozlišením se přesnost detekce sledováním pohybovala mezi 93,24 až 95,29 %. To je oproti normálnímu osvětlení pokles o 3,44 až 5,49 %. Hlavní příčinou tohoto poklesu je především častější ztráta sledovacích bodů, která je dána vlivem kompenzace osvětlení rychlostí závěrky snímače resp. prodlouženou dobou snímání jednoho snímku videa. V tomto důsledku dochází k poklesu snímků za sekundu streamovaného videa a tudíž k rozmazání obrazu ve směru pohybu hlavy. K tomuto jevu docházelo proto, že použitá kamera počítače pro snímání obrazů neumožňovala fixní nastavení expozice nebo konstantní nastavení snímků za sekundu (fps – frames per second) a tímto si kompenzovala rozložení jasových hodnot obrazu do středu histogramu. Při normálním osvětlení se video streamovalo s přibližně 20 fps (nejdelší čas snímání jednoho snímku odpovídá 1/20 sekundy) a při jiných osvětleních padalo až k 6 fps (1/6 sekundy), což je spodní hranice této kamery.

Míra spolehlivosti detekce jednotlivou detekcí je v průměru o 4,67 % nižší než při sledování obličeje. Lze tedy usoudit, že změna rozlišení má na jednotlivou detekci tváře značně větší vliv než změna osvětlení. U sledování obličeje za výše zmíněných podmínek je to právě naopak.

## 6.2 Objektivní testování robustnosti detekce a klasifikace obličejů

Tak jako každý reálný signál i obrazová data podléhají degradačním vlivům, které mohou mít původ jak ve snímané scéně, tak v technických parametrech. Testování algoritmu probíhalo podle diagramu uvedeného na Obrázek 53.



Obrázek 53: Objektivní testování robustnosti algoritmu vůči různým druhům degradace obrazu

Jedním z častých vlivů je přítomnost šumu v obraze, který může být způsobený mnoha faktory, jako je například kvalita a citlivost snímače, špatné osvětlení scény nebo také chybný přenos informace ze snímače do paměti.

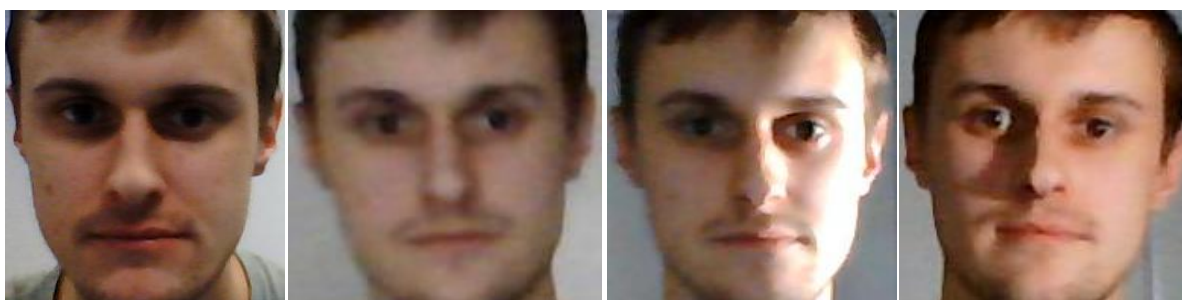
Dalším častým vlivem může být velikost pořizovaného obrazu. V dnešní době se setkáváme s velmi malými kamerami, které jsou schopny pořizovat záznam s rozlišením 4K a náklady na jejich pořízení nejsou drahé. Velmi vysoké rozlišení však není předmětem tohoto testování, protože s rostoucím rozlišením vznikají mnohonásobně vyšší požadavky na výpočetní výkon při zpracování takto obsáhlých dat. Proto bylo potřeba zjistit, při jakém nejmenším rozlišení je systém schopný správně vyhodnotit požadovaný problém a s jakou přesností.

Mezi degradační vlivy se také řadí míra expozice a osvětlení scény. Za reálných podmínek se okolní osvětlení může měnit a při manuálně nastaveném nastavení snímáče může docházet k podexponování nebo k přexponování obrazu, kdy pixely ztrácí svou informační hodnotu. Značný vliv na detekci a klasifikaci může mít také směr osvětlení scény, kdy mohou na obličejích vznikat falešné hrany v důsledku tvorby stínů.

V následujících grafech bude tmavě modrá barva představovat míru správnosti detekce obličeje. Odstíny světle modré barvy budou znázorňovat výsledky základní testovací sady DAT1 a odstíny oranžové barvy znázorňují výsledky rozšířené testovací sady dat DAT2. Tmavší barvy představují míru správně rozpoznaných obličejů a světlejší barvy představují míru důvěry klasifikace neboli skóre správně klasifikovaných obličejů.

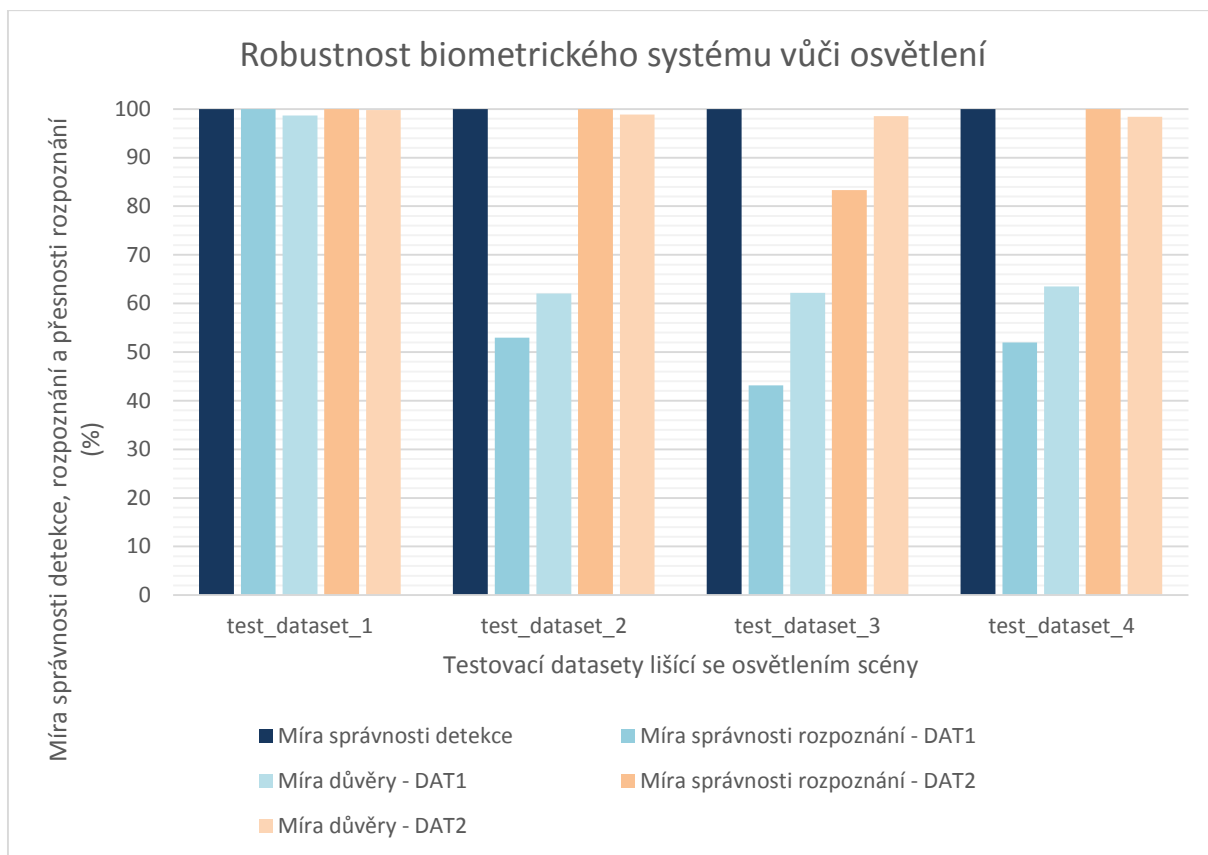
### 6.2.1 Spolehlivost detekce a klasifikace vůči změně směru osvětlení scény

Jak již bylo zmíněno, různé osvětlení a falešné hrany na obličejích mohou znehodnocovat kvalitu rozpoznání. Z tohoto důvodu byly vytvořeny celkem 4 testovací datasety o velikosti 102 obrazů, viz kapitola 5.3. Jedná se o osvětlení (popis zleva) při normálním ambientním světle (*light\_high*), při zhasnutém světle (*light\_low*), z úhlu 90° (*light\_triangle*) a z úhlu 45° (*light\_split*). Ukázka těchto osvětlení je na obrázku (Obrázek 54).



*Obrázek 54: Typy jednotlivých osvětlení zleva: test\_dataset\_1, test\_dataset\_2, test\_dataset\_3, test\_dataset\_4*

Výsledky testování vůči změně směru osvětlení scény jsou vidět v nadcházejícím grafu (Obrázek 55). Z grafu je patrné, že změna směru osvětlení nehraje žádnou roli při detekci obličeje. Naopak je tomu u samotné klasifikace. Pro trénovací sadu DAT1 je správnost klasifikace okolo 50 % a méně pro všechna různá osvětlení (tento důvod vedl k vytvoření rozšířené tréninkové databáze DAT2). Kdežto pro rozšířenou trénovací sadu DAT2 je přesnost klasifikace 100% kromě osvětlení z úhlu 90°. U všech správně rozpoznaných osob byla míra důvěry okolo 98 %.



Obrázek 55: Graf robustnosti biometrického systému vůči osvětlení

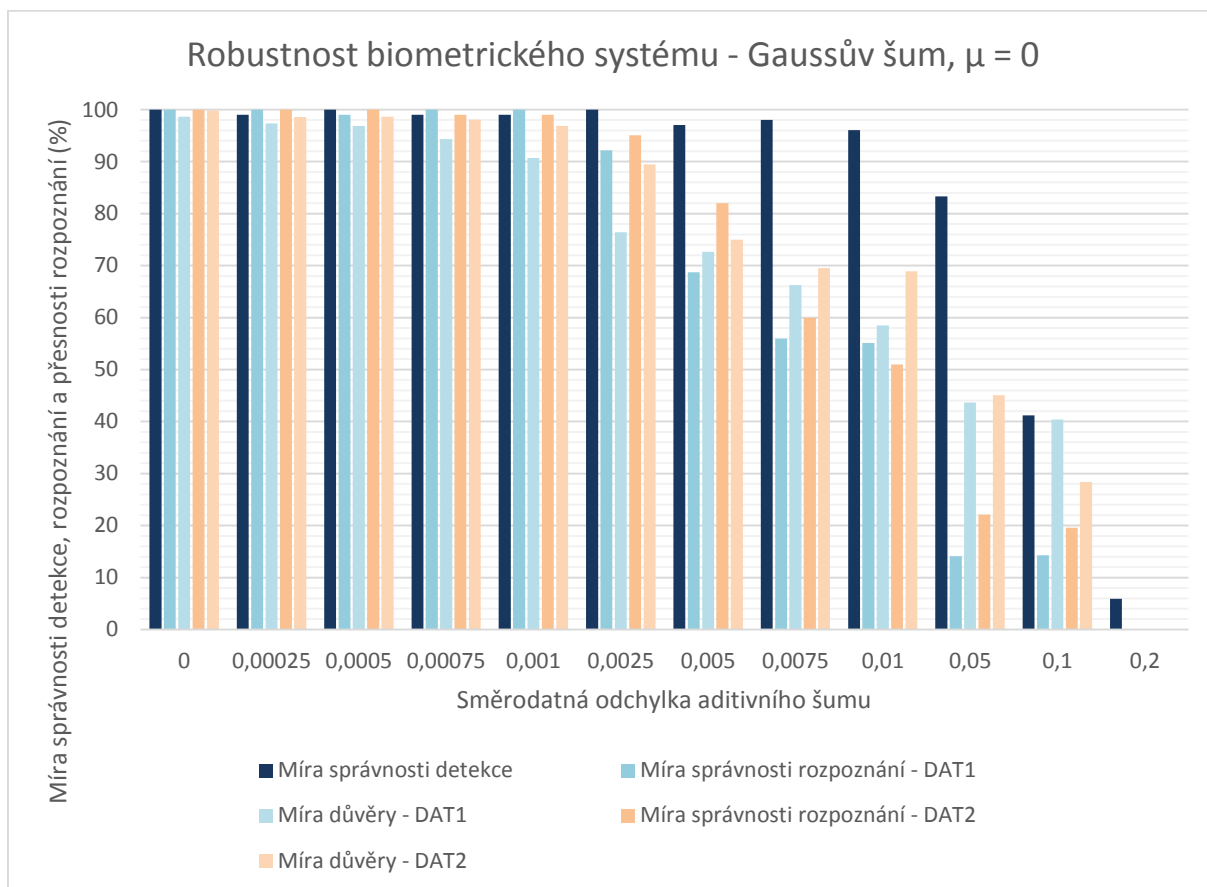
## 6.2.2 Spolehlivost detekce a klasifikace vůči přítomnosti šumu v obraze

Testování robustnosti systému probíhalo také vůči přítomnému šumu v obraze. Pro objektivní posouzení této spolehlivosti byl šum do obrazu přidáván pomocí tří různých šumových generátorů. Jednalo se o Gaussův šum, šum Salt & Pepper, a šum Speckle. Každý z těchto šumů degraduje vstupní obraz do takové míry, do jaké je nastaven jeho parametr.

- **Gaussův šum:** Gaussův aditivní bílý šum se vyznačuje normálním (Gaussovým) rozložením, kde míra šumu je na všech pixelech stejná. Tento šum se vyznačuje střední hodnotou a směrodatnou odchylkou. Při testování byla střední hodnota nulová (konstantní)  $\mu = 0$  a směrodatná odchylka proměnlivá na intervalu  $\sigma^2 = (0,00025; 0,2)$ .
- **Speckle:** Šum Speckle je multiplikativním šumem, jehož míra degradace je dána směrodatnou odchylkou, která byla při testování měněna na intervalu  $\sigma^2 = (0,00025; 0,2)$ .
- **Salt & Pepper:** Šum typu sůl a pepř je impulsní aditivní šum, který se projevuje tzv. zrněním. Parametrem tohoto šumu je jeho hustota neboli denzita v obraze, která udává procentuální zastoupení zašuměných pixelů v obraze. Rozsah hustoty tohoto šumu byl na intervalu  $(0,00025; 0,2)$ .

Odolnost algoritmu vůči Gaussovu šumu můžeme vidět na tomto grafu (Obrázek 56). Míra správnosti detekovaných obličejů klesá pod 95 % až při hodnotě směrodatné odchylky 0,01, z čehož lze usoudit,

že kvalita detekce je na velmi dobré úrovni. To stejné se však, ale nedá říct u správnosti klasifikace, která při směrodatné odchylce 0,005 je už jen mezi 68 až 82 %. Pro obě tyto tréninkové sady jsou výsledky při zašumění Gaussovým šumem téměř podobné. Ukázka vstupního snímku se dvěma hodnotami směrodatné odchylky Gaussova šumu je znázorněna na obrázku (Obrázek 57).



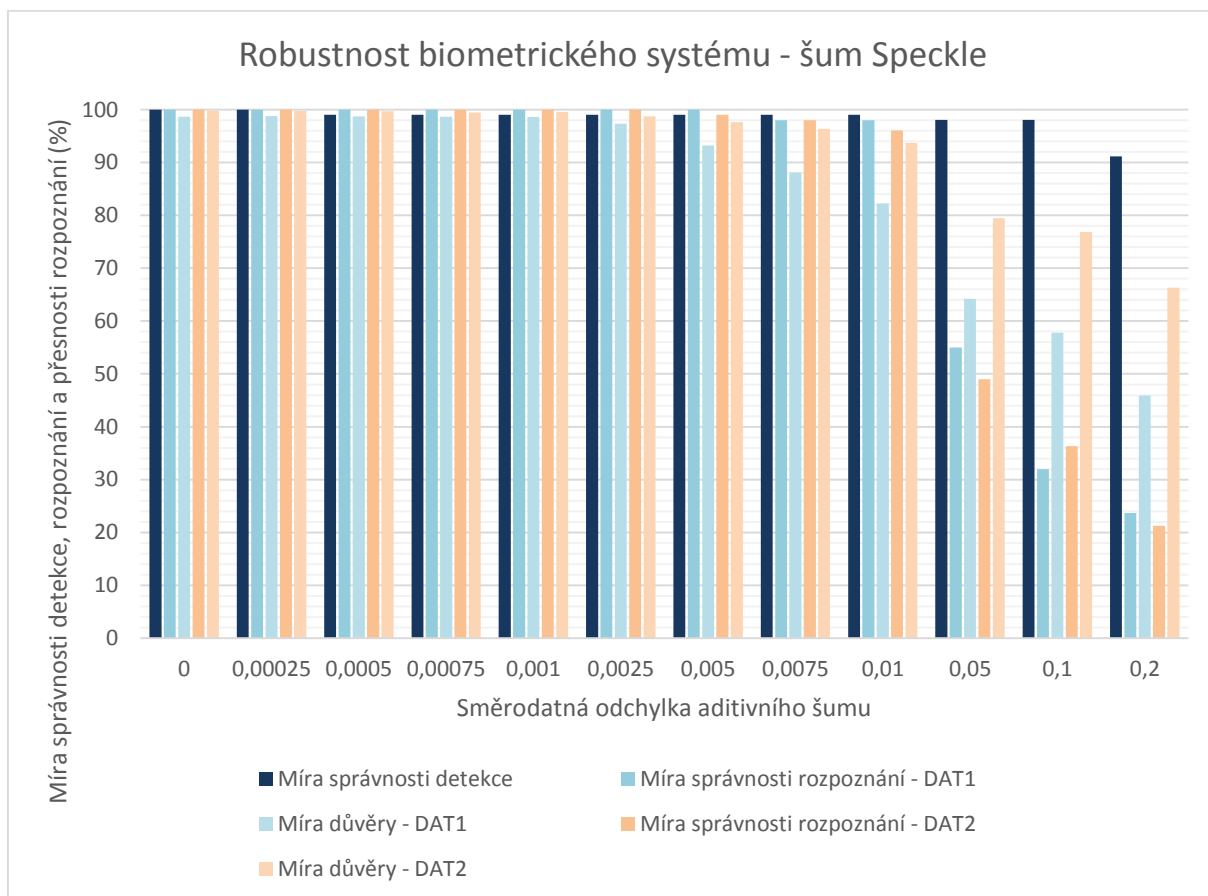
Obrázek 56: Graf robustnosti biometrického systému vůči změně směrodatné odchylky Gaussova šumu



Obrázek 57: Ukázka degradace vstupního obrazu Gaussovým šumem

Ve srovnání s odolností systému na šum typu Speckle (Obrázek 58) je vidět, že při stejném rozsahu směrodatné odchylky správnost detekce obličeje neklesá pod 90 %. Robustnost klasifikace vůči tomuto šumu je rovněž větší než v případě Gaussova šumu, kdy do směrodatné odchylky 0,01 neklesá správnost

rozpoznání pod 96 % (Obrázek 59). V tomto grafu si lze také všimnout, že mezi základní a rozšířenou trénovací databází je rozdíl v míře důvěry rozpoznaných obličejů. Nejlépe je tato situace vidět při směrodatné odchylce 0,005, 0,0075 a 0,01, kde přesnost rozpoznání obou datových sad je téměř stejná, avšak přesnost rozhodování byla u rozšířené sady o 4, 10 a 12 % vyšší.



Obrázek 58: Graf robustnosti biometrického systému vůči změně směrodatné odchylky šumu Speckle



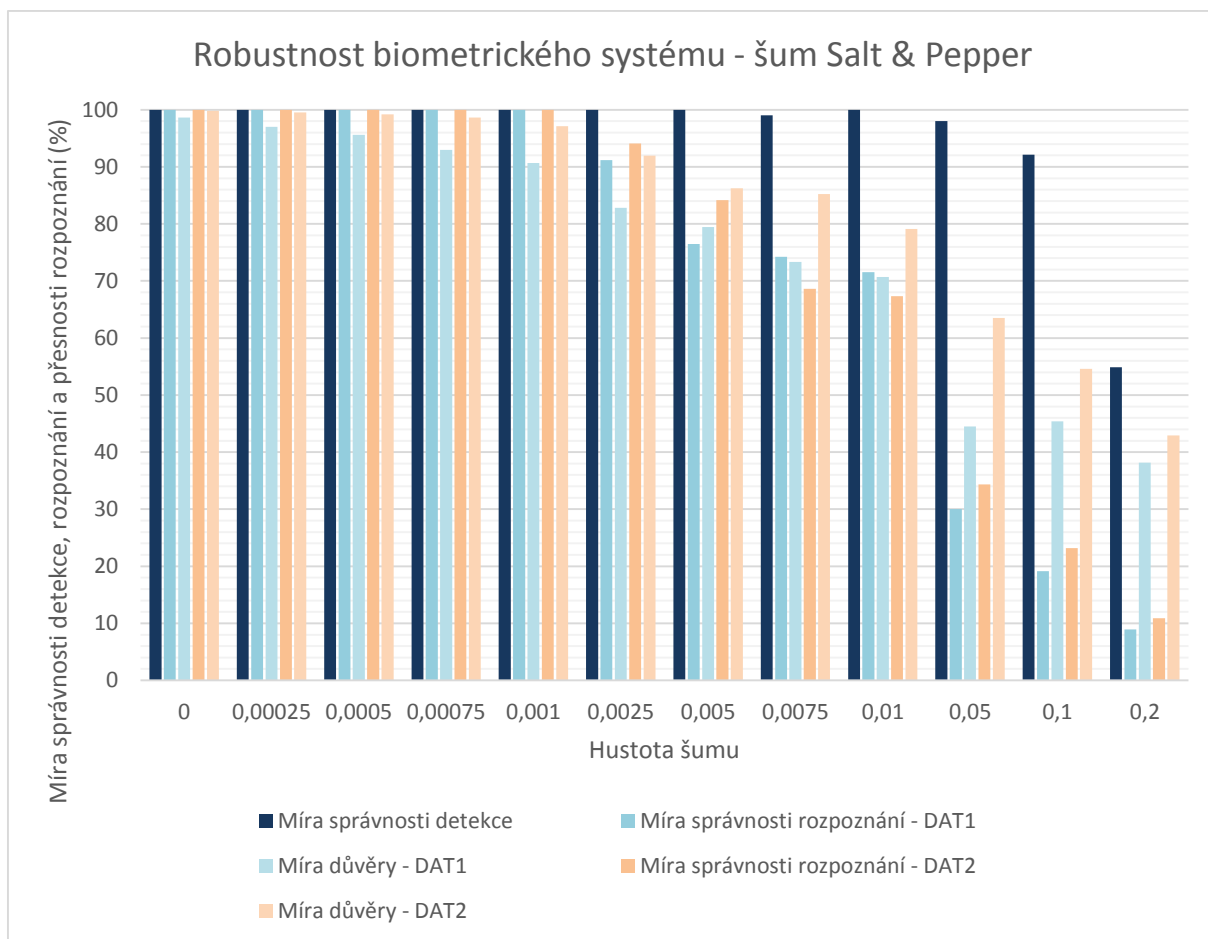
Obrázek 59: Ukázka degradace vstupního obrazu šumem Speckle

Testování vůči hustotě šumu Salt and Pepper je znázorněno grafem na obrázku (Obrázek 60). Obraz ovlivněný hustotou tohoto šumu do míry 0,05 vykazuje správnost detekce 98 %. V následujících dvou



krocích je vidět největší propad spolehlivosti detekce. Na druhou stranu při hodnotě 0,2 je pokryto šumem 20 % obrazu, z čehož vyplývá, že takovýto pokles je očekávaný (Obrázek 61).

Při hodnocení klasifikace s přibývajícím šumem do hustoty 0,001 postupně klesá skóre rozpoznaných obličejů, které byly do této chvíle všechny rozpoznány správně. Při hustotě šumu 0,0025 začíná rapidně klesat míra správnosti klasifikace obličejů.



Obrázek 60: Graf robustnosti biometrického systému vůči změně denzity šumu Salt & Pepper



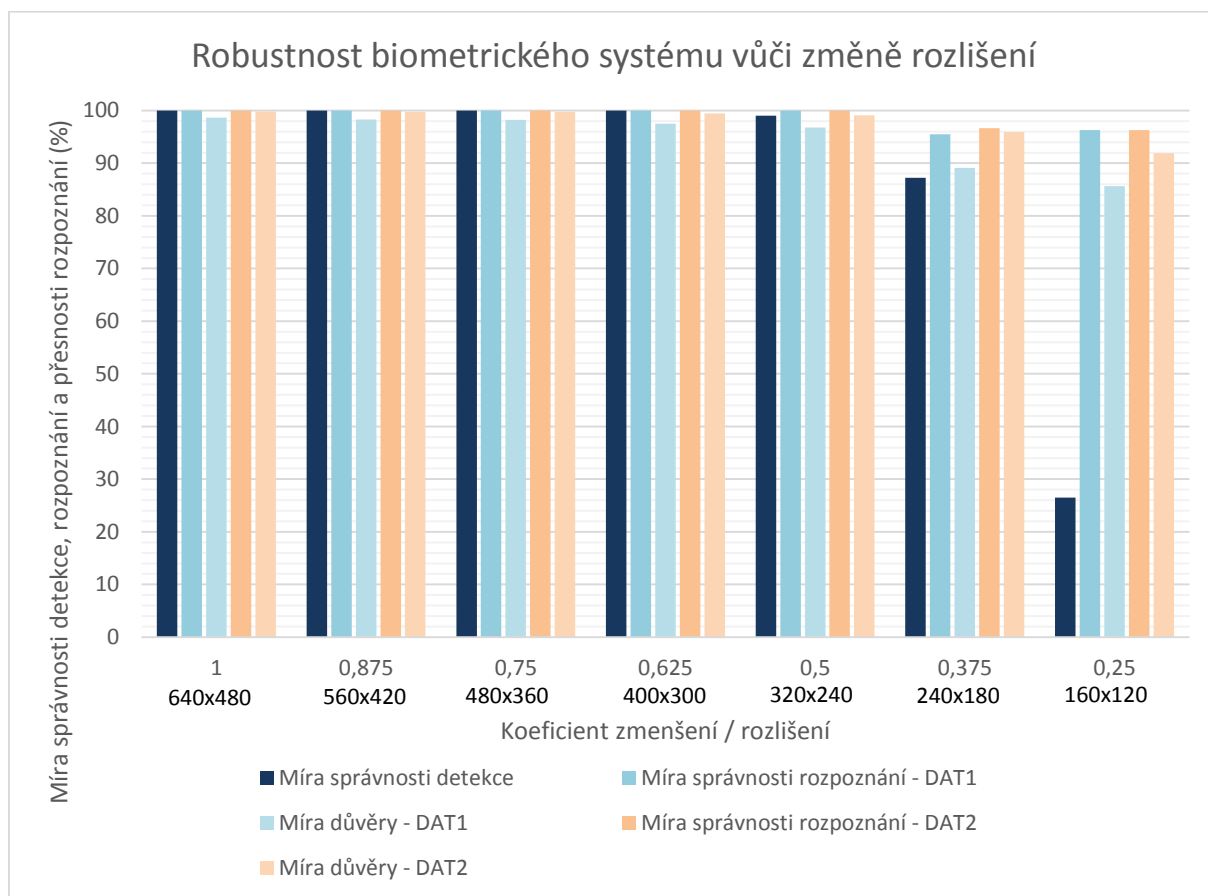
Obrázek 61: Ukázka degradace vstupního obrazu šumem Salt & Pepper

Na všech grafech hodnotících přesnost klasifikace na základě degradace obrazu šumem je vidět, že rozšířená datová sada DAT2 vykazuje, až na pár výjimek, lepší výsledky než sada základní DAT1. Pro biometrický systém je tedy výhodnější využívat rozšířenou datovou sadu DAT2. Nevýhodou této datové sady je potřeba zajistit dostatečný počet trénovacích obrazů za různých druhů osvětlení a doba trénování sítě z této databáze zabere značně delší čas.

### 6.2.3 Spolehlivost detekce a klasifikace vůči změně rozlišení obrazu

Při testování spolehlivosti algoritmu na základě změny rozlišení, byla provedena úprava ve zdrojovém kódu, kde je nyní možné detekovat obličej při minimální velikosti 60x60 pixelů. Podvzorkování obrazu bylo provedeno v 6 krocích až do čtyřnásobného zmenšení.

Úplně nalevo je sloupec (Obrázek 62) s původním obrazem s nejvyšším rozlišením a napravo pak obrazy s menší velikostí. Při změně velikosti testovacích obrazů na polovinu tj. 320x240 pixelů byla správná detekce obličejů okolo 99 % a klasifikace pro obě datové sady 100%. Při zmenšení rozlišení na čtvrtinu byla správnost detekce pouze 26 % a to z důvodu, že se osoby při tvorbě databáze pohybovaly směrem dopředu a dozadu směrem od kamery, která je snímala a tudíž v takto malém rozlišení už nebyl obličej v obraze větší než 60x60 pixelů. Přesnost správně rozpoznaných obličejů v žádném z těchto případů nespadla pod 95 %. Jako u předchozích testování je vidět, že rozšířená datová sada DAT2 klasifikuje s vyšší přesností, avšak není zde vidět rapidní rozdíl mezi mírou správnosti klasifikovaných obličejů.

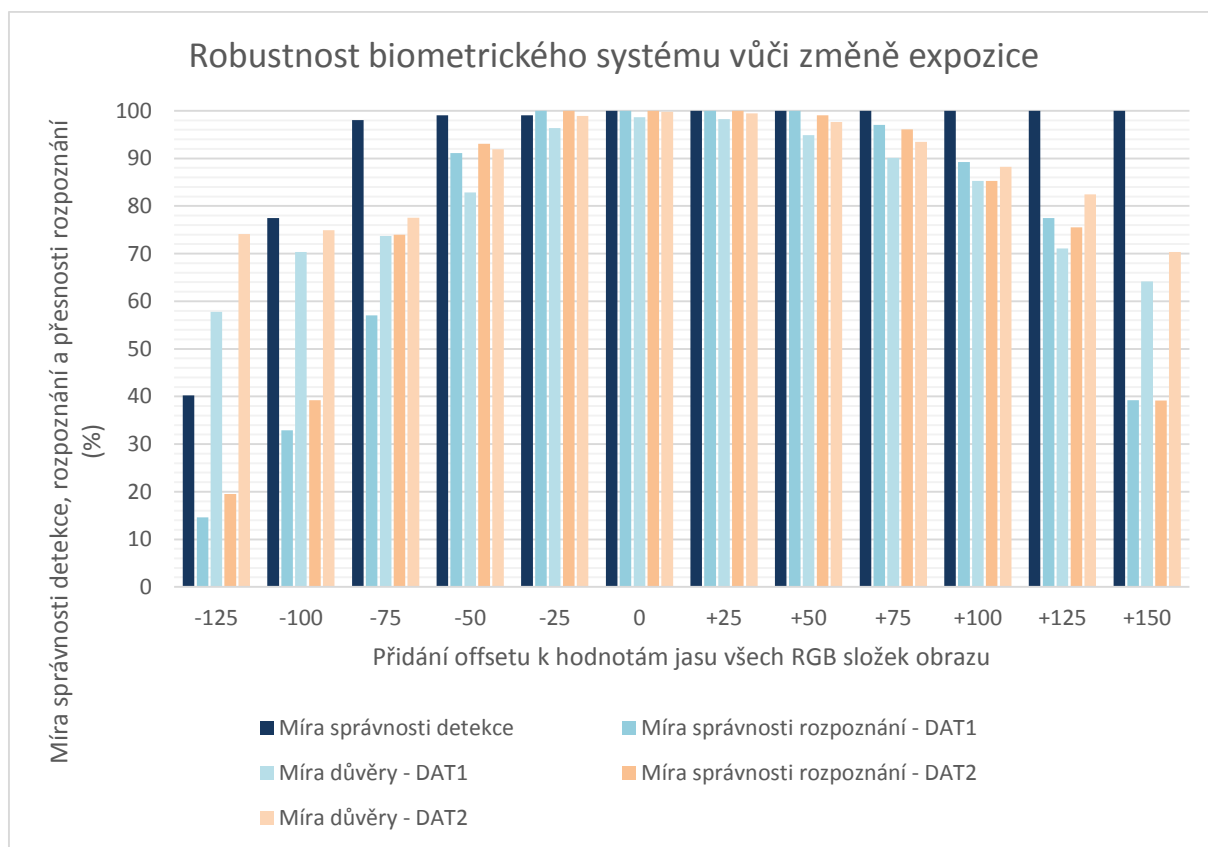


Obrázek 62: Graf robustnosti biometrického systému vůči změně rozlišení testovaného obrazu

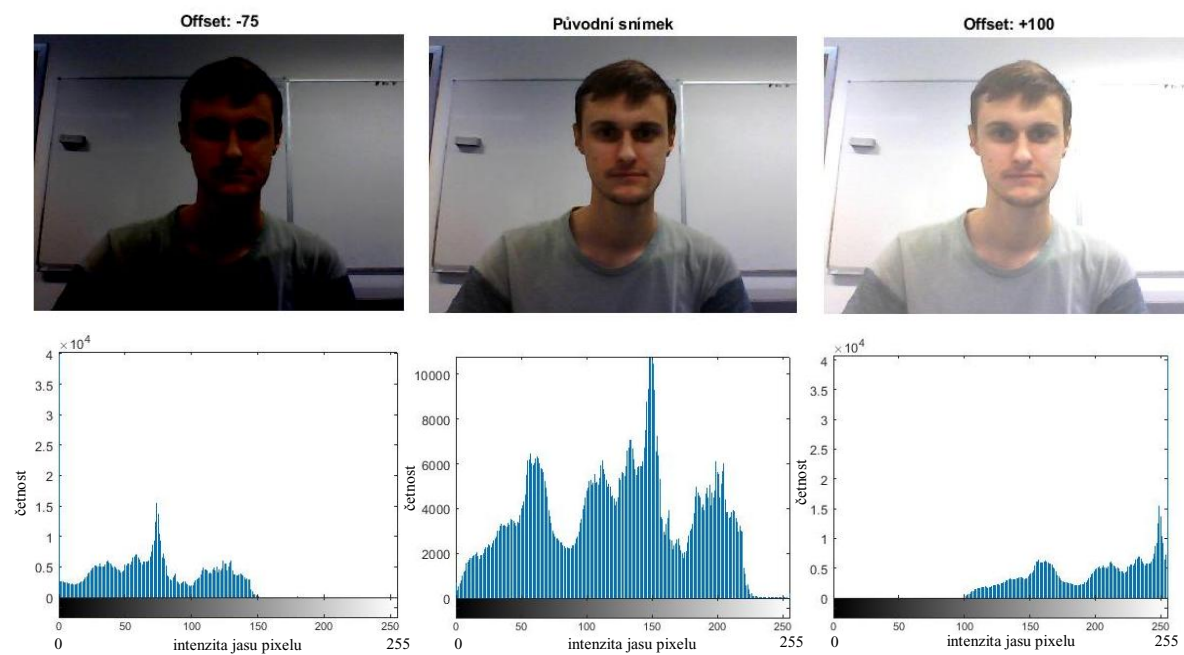


#### 6.2.4 Spolehlivost detekce a klasifikace vůči změně expozice obrazu

Poslední testování systému v této práci bylo provedeno vůči změně expozice obrazu. Expozice byla upravována přidáváním offsetu jasových hodnot k testovaným obrazům (Obrázek 64). Z grafu (Obrázek 63) je patrné že míra správnosti detekce se drží nad 98 % s výjimkou prvních dvou sloupců, kdy v těchto případech muselo docházet ke značnému podexponování obličejů. Při míře správnosti rozpoznání je vidět rozdíl mezi jednotlivými trénovacími sadami DAT1 a DAT2. V případě podexponování je na tom lépe rozšířená sada. Naopak tomu je při zvyšování expozice, kdy s rostoucím offsetem je lépe schopna klasifikovat obličeje základní trénovací sada. Zvláštní je, že v celém tomto testování je skóre pro rozpoznání obličeje vždy lepší u rozšířené sady DAT2, i přesto že míra správnosti klasifikace je nižší.



Obrázek 63: Graf robustnosti biometrického systému vůči změně expozice testovaného obrazu



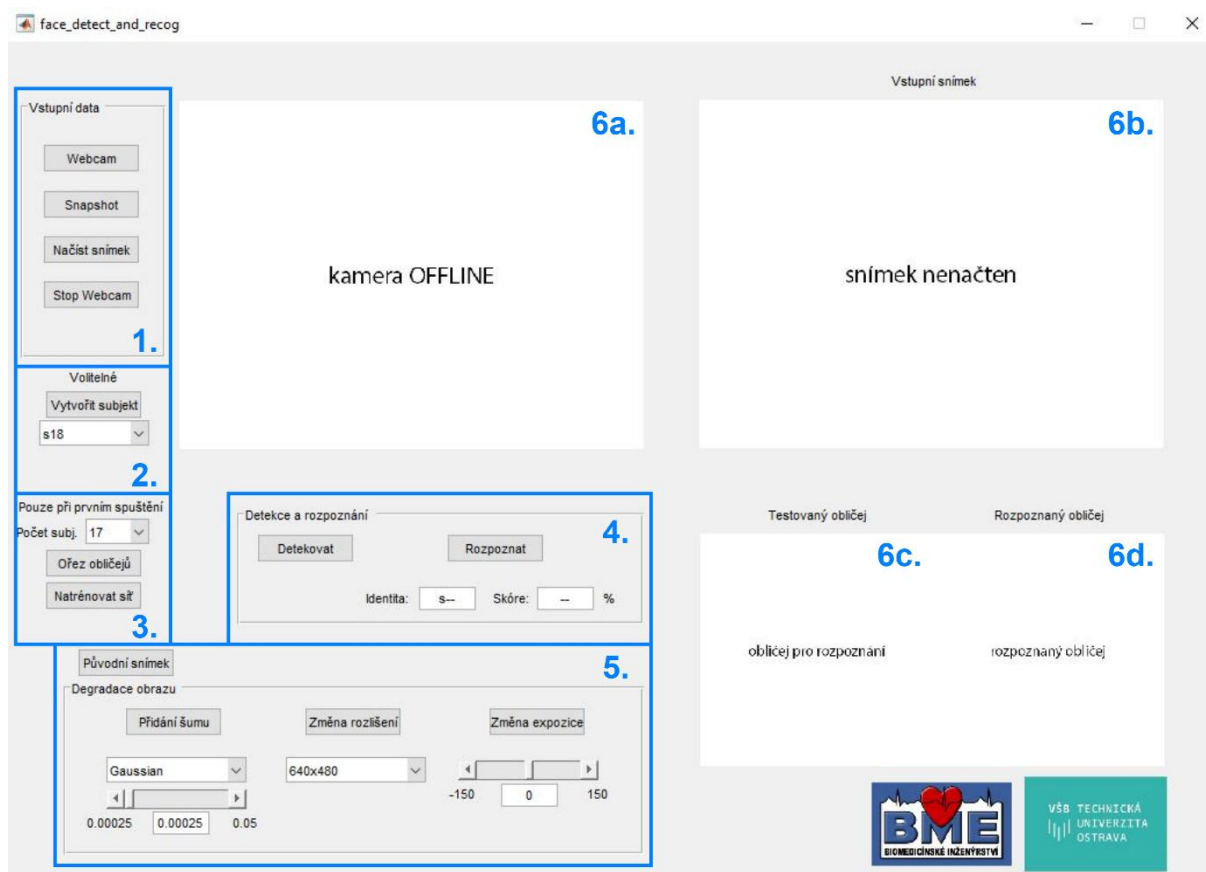
Obrázek 64: Ukázka změny expozice obrazu offsetem

## 7 Realizace SW prostředí pro biometrickou identifikaci uživatele

V poslední části práce bylo vytvořeno softwarové prostředí pro biometrickou identifikaci uživatele a simulaci variabilních podmínek pro detekci a klasifikaci obličeje. Grafické uživatelské rozhraní (GUI - Graphical User Interface) také umožňuje tvorbu nového subjektu a jeho začlenění do neuronové sítě.

Základní rozložení SW prostředí (Obrázek 65) se člení na 6 částí.

1. Panel pro načtení vstupních dat
2. Oblast volitelného nastavení, zahrnující tvorbu nového subjektu do databáze
3. Extrakce obličejů a natrénování sítě
4. Detekce a rozpoznání obličeje
5. Degradace testovaného obrazu
6. Okna pro vykreslení snímků a), b), c), d)

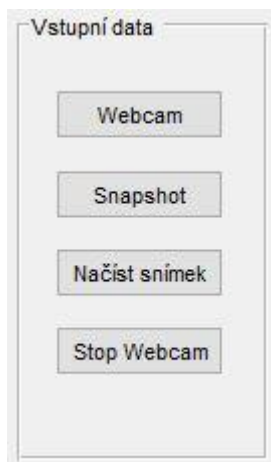


Obrázek 65: Design softwarového prostředí ve stavu idle s popisem

### 7.1 Panel pro načtení vstupních dat

Panel pro načtení vstupních dat obsahuje 4 tlačítka (Obrázek 66). Tlačítko s názvem „Webcam“ spouští živý náhled skrze kameru počítače do okna 6a (Obrázek 65). Snímek pro testování lze načíst dvěma způsoby. Jedná se pořízení aktuálního snímku z právě zmíněného videopřenosu tlačítkem „Snapshot“ nebo načtením snímku z adresáře v počítači tlačítkem „Načíst snímek“. Snímek určený pro testování se

zobrazí v okně 6b (Obrázek 65) včetně os udávajících rozlišení obrazu. Pro ukončení živého videopřenosu slouží poslední tlačítko v tomto panelu.



*Obrázek 66: Panel pro načtení vstupních dat*

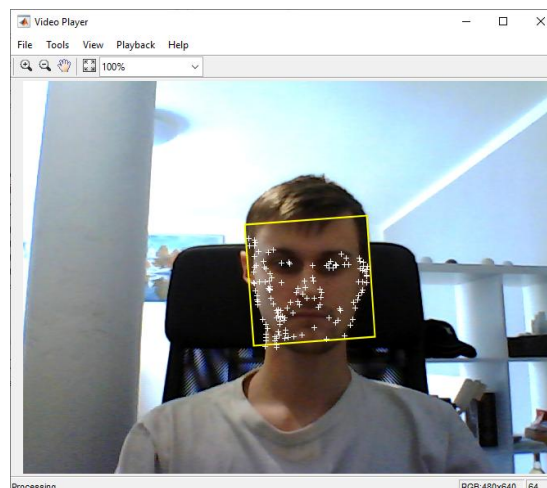
## 7.2 Oblast volitelného nastavení

Oblast volitelného nastavení obsahuje jedno tlačítko a jedno vyskakovací menu s názvem nového subjektu, který chceme přidat do databáze (Obrázek 67). V tomto případě je potřeba postupovat podle pořadí osob v databázi (defaultně je nastaven nový subjekt pod názvem „s18“).



*Obrázek 67: Přidání nového subjektu do databáze*

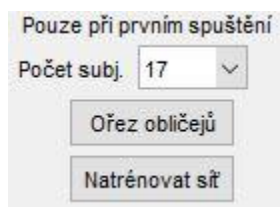
V tuto chvíli lze vytvořit, trackovací metodou, snímky dané osoby pomocí kamery počítače. Videosekvence se zobrazí v novém okně (Obrázek 68) a po dobu 100 snímků hledá obličej v obraze, které ukládá do příslušné složky v počítači („photos\_1“).



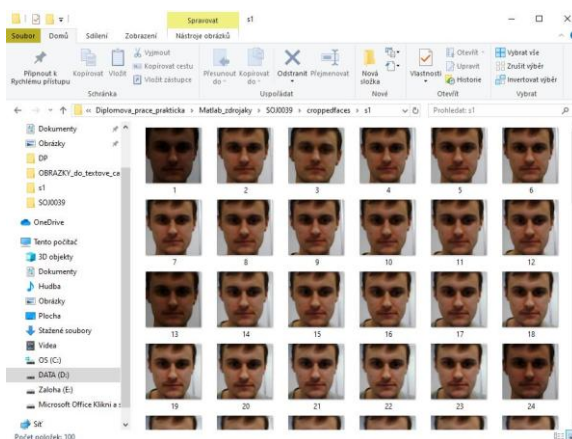
Obrázek 68: Videosekvence pořizující snímky nového subjektu

### 7.3 Extrakce obličejů a natrénování sítě

Před natrénováním konvoluční neuronové sítě je potřeba zvolit počet subjektů, pro které budeme síť trénovat. V tomto případě bude trénink probíhat ze 17 osob (důvodem výběru 17 osob je ten, že má osoba je již v databázi na první pozici a objevovala by se zde podruhé). V této chvíli je potřeba detekovat ze základní databáze obličeje a normalizovat je podle vstupu do sítě. K tomuto účelu slouží tlačítko „Ořez obličejů“ (Obrázek 69), které projde všechny snímky v databázi, detekuje obličeje a uloží je do příslušné složky v „croppedfaces\_1“ (Obrázek 70).



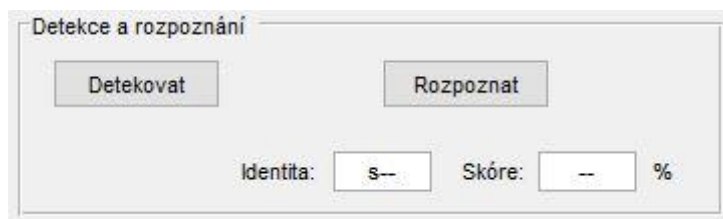
Obrázek 69: Volba počtu subjektů, ořez obličejů a natrénování sítě



Obrázek 70: Nově vytvořené a normalizované snímky v příslušné složce databáze

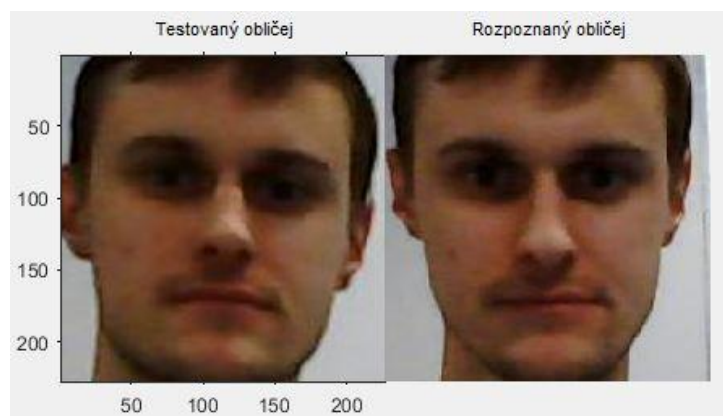
## 7.4 Detekce a rozpoznání

Panel pro detekování a rozpoznání obličejů se skládá ze dvou stejnojmenných tlačítek (Obrázek 71). Po úspěšné detekci obličeje se normalizovaný snímek načte do okna 6c (Obrázek 65). Nyní je možné určit identitu osoby, jejíž jméno („s1“, „s2“, atd.) se zobrazí v prvním textovém poli. Do druhého textového pole je zapsána míra důvěry klasifikace v procentech.



Obrázek 71: Oblast programu pro detekci a klasifikaci testovaného obrazu

Ve stejnou chvíli se pro ilustraci také zobrazí identifikovaná osoba v okně 6d (Obrázek 65 a Obrázek 72).

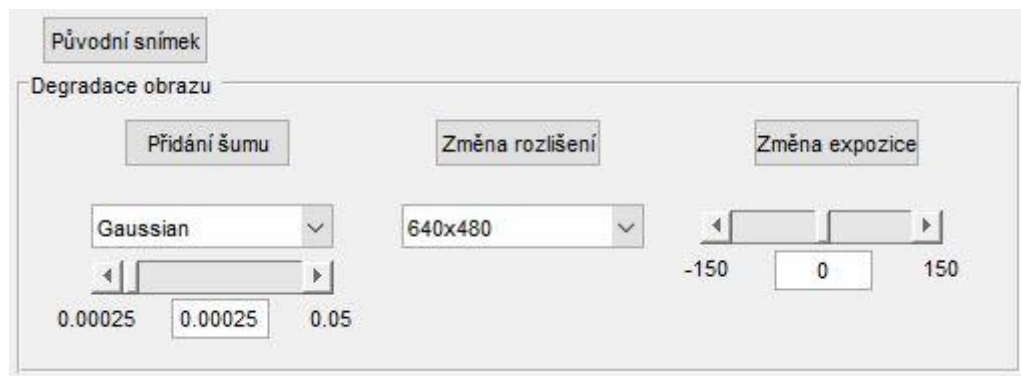


Obrázek 72: Testovaný obličej a obličej identifikované osoby

## 7.5 Degradace testovaného obrazu

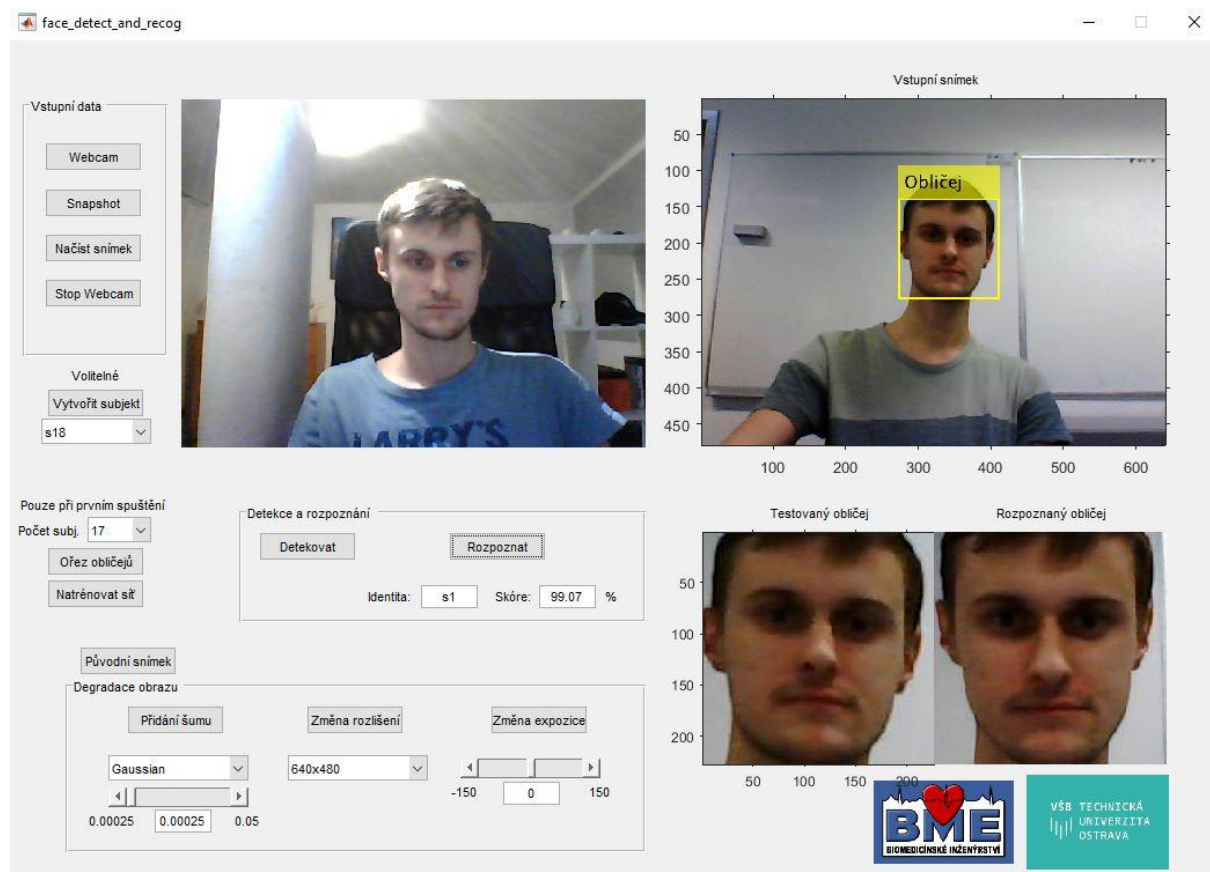
Načtený snímek v okně 6b (Obrázek 65) je možné degradovat přidáním šumu, změnou velikosti a změnou expozice. Všechny aplikace lze najít v oblasti degradace testovaného obrazu (Obrázek 73). Jako první je potřeba vybrat z vyskakovacích oken příslušný šum nebo velikost rozlišení. Dalším krokem je nastavení parametrů dané aplikace pomocí tzv. slideru, kterým se upravuje intenzita šumu nebo offsetu expozice. Poté je možné aplikovat danou degradaci stiskem příslušného ovládacího prvku a změna se projeví v již zmíněném okně 6b.

Pokud je potřeba získat původní testovaný snímek bez degradačních vlivů, lze jej obnovit stiskem tlačítka „Původní snímek“.



Obrázek 73: Sekce SW prostředí pro degradaci testovaného snímku

Ukázka softwarového prostředí za běhu je znázorněna následujícím obrázkem (Obrázek 74).



Obrázek 74: Ukázka SW prostředí v průběhu používání aplikace



## 8 Zhodnocení výsledků práce

Cílem této práce bylo realizovat biometrický systém, který bude schopen rozpoznávat identitu osob podle faciálních rysů člověka. Součástí práce bylo také objektivní vyhodnocení robustnosti a efektivity realizovaného biometrického systému na základě stanovených obrazových degradačních aplikací.

Teoretická část shrnuje pojem biometriky a jeho nedílných součástí, základní operace pro zpracování obrazu a rešerši metod pro detekci, sledování a klasifikaci obličejů z obrazových dat.

Praktická část práce se skládala z několika základních bodů. Pro účely trénování a testování algoritmů byla vytvořena vlastní databáze obličejů se 17 jedinci a s celkovým počtem 10540 obrazů za různých podmínek. Tvorba této databáze byla uskutečněna za pomoci kamery počítače, která ukládala ze živého videopřenosu obrazy s detekovaným obličejem na úložiště na základě sledovacího algoritmu KLT. Z testování bylo zjištěno, že sledovací algoritmus dokáže detekovat obličej v obraze, při rozlišení 640x480 a za normálního světla, s mírou přesnosti 98,73 %. Oproti tomu, klasická detekce kaskádní metodou Viola-Jones z jednotlivých obrazů byla přesná pouze s 92,75 %. Za všech variabilních podmínek dokázal sledovací algoritmus správně detekovat tváře o minimálně 3,24 % lépe, než jednotlivá detekce kaskádním klasifikátorem. Veškeré výsledky jsou uvedeny v tabulce (Tabulka 3).

V další části práce bylo využito předtrénované konvoluční neuronové sítě AlexNet pro klasifikaci obličejů. Neuronová síť byla naučena pomocí dvou databází, tj. základní DAT1 a rozšířenou DAT2. Tyto dvě sítě byly následně otestovány vůči různým degradačním vlivům, mezi které patří změna směru osvětlení scény, 3 generátory šumu (Gaussův, Speckle, Sůl a Pepř), změna rozlišení a změna expozice.

V reakci na změnu směru osvětlení scény si mnohem lépe vedla síť naučená rozšířenou datovou sadou DAT2. Při kvantitativní komparativní analýze degradace obrazu šumovými generátory, vykazoval algoritmus nejlepších výsledků při šumu Speckle, kde přesnost detekovaných obličejů nesestoupila pod 90 % (na celém rozsahu testovaných hodnot) a kvalita klasifikace byla do velikosti směrodatné odchylky 0,01 vyhodnocena s přesností nad 95 %. S narůstající hodnotou směrodatné odchylky Gaussova šumu a hustoty šumu Salt & Pepper nad hodnotu 0,0025 lze sledovat rapidní pokles přesnosti klasifikace. Do hodnoty směrodatné odchylky a hustoty 0,001 se pohybovala přesnost detekce a rozpoznání na 98 %.

Testování systému v závislosti na velikosti rozlišení vstupního obrazu mělo vliv především na míru správně detekovaných obličejů, jehož přesnost při 4 násobném zmenšení (160x120) obrazu byla pouhých 26,47 %. Při změně rozlišení na polovinu, tedy 320x240, dosahovala tato přesnost hodnoty minimálně 99,01 %. Klasifikace osob při polovičním rozlišení dosahovala správnosti rozpoznání 100 %. I přes změnu rozlišení na čtvrtinu byla přesnost klasifikace minimálně 95,5 % pro obě sítě. Robustnost systému byla také testována vůči změně expozice snímku, kde při postupném ztmavování obrazu klesala především správnost klasifikace sítě, natrénované základní databází DAT1. V opačném případě, při zvyšování expozice testovaného snímku docházelo k většímu poklesu správnosti klasifikace u sítě natrénované rozšířenou datovou sadou DAT2.

Veškerá testování s podrobnějším popisem efektivity systému jsou uvedena v kapitole 6 a v přílohách.

Pro výslednou reprezentaci systému bylo vytvořeno SW prostředí (kapitola 7), které umožňuje jednoduše otestovat implementované algoritmy. Program umožňuje veškeré procesy popsané v této práci, tj. od samotného vytvoření subjektu, přes degradaci testovaného snímku, až po konečnou klasifikaci uživatele s výpisem míry důvěry klasifikace.



## **Závěr**

V této diplomové práci byl realizován unimodální biometrický systém pro autentizaci osob na základě obličeje ve vývojovém prostředí MATLAB. Systém je schopný detekovat obličeje ze statických obrazů a sledovat je v dynamických obrazech. Při práci s dynamickými obrazy byl kladen požadavek na rychlost zpracování dat a z tohoto důvodu byla při realizaci systému zvolena metoda detekce obličeje Viola-Jones a metoda sledování obličeje KLT. Pro identifikaci osob byla implementována metoda hlubokých neuronových sítí na principu konvoluce (CNN – konvoluční neuronová síť). Součástí práce byla tvorba databáze faciálních obrazů, návrh a realizace biometrického systému, testování robustnosti a efektivity algoritmů a tvorba uživatelského SW prostředí pro klasifikaci osob.

Vlastní databáze faciálních obrazů byla pořízena kamerou počítače za různých podmínek pro 17 subjektů o celkovém počtu 10 540 obrazů. Databáze byla následně rozdělena na dvě trénovací databáze a na databáze sloužící pro testování. Základní trénovací databáze obsahovala 100 snímků na subjekt při stejných světelných podmínkách. Rozšířená trénovací databáze obsahovala 250 snímků na subjekt, kde k základní databázi bylo přidáno 3x50 snímků při různých osvětleních snímané scény. Rozdíl, mezi natrénováním CNN jednotlivými databázemi, byl především v přesnosti výsledné klasifikace, ale také ve výpočetním čase pro její natrénování. Čas potřebný pro natrénování sítě základní databází DAT1 byl přibližně 11 minut, kdežto doba trénování sítě rozšířenou databází DAT2 trvala cca 21 minut.

V průběhu práce byla srovnána míra správnosti detekovaných obličejů sledovacím algoritmem a jednotlivou detekcí obličeje. Dále pak byl systém objektivně testován na testovacím datasetu, který procházel různými formami degradace. Testovací obrazy byly postupně znehodnocovány třemi druhy aditivních šumů, změnou rozlišení a změnou expozice. Systém byl rovněž otestován na datasetech s různým osvětlením scény. Detailní výsledky robustnosti systému lze nalézt v kapitole 6 a 8.

Výslednou realizaci systému reprezentuje vytvořené SW prostředí pro biometrickou klasifikaci uživatele na základě rozpoznání obličeje. Program umožňuje vytvoření nového subjektu do databáze, načtení testovacího snímku z disku počítače nebo ze živého náhledu prostřednictvím kamery počítače a natrénování konvoluční neuronové sítě. Dále pak umožňuje z testovacího obrazu detekovat a rozpoznat obličej vzhledem k natrénované síti a určit přesnost jeho predikce. Vstupní testovací obraz může být také podroben degradaci již zmíněnými degradačními metodami s okamžitým zobrazením znehodnoceného snímku v okně programu. Detailní popis SW prostředí lze nalézt v kapitole 7.

Podrobná data o testování jsou k dispozici v textu práce, v přílohách nebo v přiložených souborech v elektronické podobě na CD. Elektronická příloha také obsahuje návod ke spuštění a k chodu programu.

## Zdroje

- [1] DRAHANSKÝ, Martin, Michal DOLEŽEL a Filip ORSÁG. *Biometrie*. Brno: [M. Drahanský], 2011. ISBN 978-80-254-8979-6.
- [2] DRAHANSKY, Martin. Liveness Detection in Biometrics. *Advanced Biometric Technologies* [online]. InTech, 2011 [cit. 2017-02-14]. DOI: 10.5772/17205. ISBN 978-953-307-487-0
- [3] DOLEŽEL, Michal. *Falsifikace biometrické charakteristiky a detekce zivosti* [online]. 20.4.2014, 43 [cit. 2020-01-03]. Dostupné z: <https://www.slideshare.net/securitysession/falsifikace-biometricke-charakteristiky-a-detekce-zivosti>
- [4] Voice recognition (speaker recognition). *TechTarget: Search Customer Experience* [online]. 2018 [cit. 2020-01-03]. Dostupné z: <https://searchcustomerexperience.techtarget.com/definition/voice-recognition-speaker-recognition>
- [5] ŠČUREK, Radomír. *Biometrické metody identifikace osob v bezpečnostní praxi* [online]. VŠB TU Ostrava, Červen 2018, 58 [cit. 2020-01-03]. Dostupné z: [http://www.rucnepsanypodpis.cz/PDF/biometricke\\_metody.pdf](http://www.rucnepsanypodpis.cz/PDF/biometricke_metody.pdf)
- [6] ROSS, A.; JAIN, A.: Multimodal biometrics: an overview. [online]. 2004, [cit. 2020-01-03]. Dostupné z: [http://www.cse.msu.edu/rgroups/biometrics/Publications/Multibiometrics/RossJain\\_MultimodalOverview\\_EUSIPCO04.pdf](http://www.cse.msu.edu/rgroups/biometrics/Publications/Multibiometrics/RossJain_MultimodalOverview_EUSIPCO04.pdf)
- [7] BOLLE, Ruud M. et.al. *Guide to Biometrics*. New York: Springer Science & Business Media, 2004. ISBN 0-387-40059-3
- [8] *WikiSkripta: Digitální obraz* [online]. 2016 [cit. 2020-01-03]. Dostupné z: [https://www.wikiskripta.eu/w/Digit%C3%A1ln%C3%AD\\_obraz](https://www.wikiskripta.eu/w/Digit%C3%A1ln%C3%AD_obraz)
- [9] ROHRER, Brandon. *How to Convert an RGB Image to Grayscale: End-to-End Machine Learning library* [online]. November 14, 2019 [cit. 2020-01-03]. Dostupné z: [https://brohrer.github.io/convert\\_rgb\\_to\\_grayscale.html](https://brohrer.github.io/convert_rgb_to_grayscale.html)
- [10] HLAVÁČ, Václav. *Jasové a geometrické transformace* [online]. ČVUT Praha, , 37 [cit. 2020-01-03]. Dostupné z: <http://people.ciirc.cvut.cz/~hlavac/TeachPresCz/11DigZprObr/18BrightGeomTxCz.pdf>
- [11] HLAVÁČ, Václav. *Hledání hran a hranových bodů* [online]. ČVUT Praha, , 47 [cit. 2020-01-03]. Dostupné z: <http://people.ciirc.cvut.cz/~hlavac/TeachPresCz/11DigZprObr/22EdgeDetectionCz.pdf>
- [12] JANÁKOVÁ, Ilona. *Computer Vision Group: Segmentace* [online]. VUT Brno, , 31 [cit. 2020-01-03]. Dostupné z: [http://midas.uamt.feec.vutbr.cz/POV/Lectures/05\\_Segmentace.pdf](http://midas.uamt.feec.vutbr.cz/POV/Lectures/05_Segmentace.pdf)
- [13] MING-HSUAN, Yang. *Detecting Faces in Images: A Survey* [online]. IEEE, January 2002, , 25 [cit. 2020-01-03]. Dostupné z: <https://vision.cornell.edu/se3/wp-content/uploads/2014/09/pami02.pdf>
- [14] KONTROPOULOS, Constantine a Ioannis PITAS. *Rule-based Face Detection in Frontal Views*. Greece, 2016. Aristotle University of Thessaloniki
- [15] KOLKUR, S. a D. KALBANDE. *Human Skin Detection Using RGB, HSV and YCbCr Color Models* [online]. Atlantis Press, 2017, , 9 [cit. 2020-01-03]. Dostupné z: <https://arxiv.org/ftp/arxiv/papers/1708/1708.02694.pdf>
- [16] YCbCr. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2019 [cit. 2020-01-03]. Dostupné z: <https://en.wikipedia.org/wiki/YCbCr>
- [17] Berbar, Mohamed & Kelash, Hamdy & Kandeel, Amany. (2006). *Faces and Facial Features Detection in Color Images*. 2006. 209-214. 10.1109/GMAI.2006.18. Dostupné z: [https://www.researchgate.net/publication/221414213\\_Faces\\_and\\_Facial\\_Features\\_Detection\\_in\\_Color\\_Images](https://www.researchgate.net/publication/221414213_Faces_and_Facial_Features_Detection_in_Color_Images)

- [18] PONZER, Martin. *Detekce a Rozpoznávání obličejů* [online]. Brno, 2009 [cit. 2020-01-03]. Diplomová práce. VUT Brno. Dostupné z: [https://www.vutbr.cz/studenti/zav-prace?zp\\_id=22226&aid\\_redir=1](https://www.vutbr.cz/studenti/zav-prace?zp_id=22226&aid_redir=1)
- [19] HJELMAS, Erik. *Face Detection: A Survey* [online]. Academic Press, 2001, , 39 [cit. 2020-01-03]. Dostupné z: <http://groups.di.unipi.it/~cardillo/RN1/papers/hjelmas01face.pdf>
- [20] Yuille, A.L., Hallinan, P.W., & Cohen, D.S. (1989). *Feature extraction from faces using deformable templates*. *International Journal of Computer Vision*, 8, 99-111. Dostupné z: <https://www.semanticscholar.org/paper/Feature-extraction-from-faces-using-deformable-Yuille-Hallinan/df59d8c95025bcc540b31e04fa2d3a702404fb6f>
- [21] Bukis, Audrius & Proscavicius, Tomas & Raudonis, Vidas & Simutis, Rimvydas. (2011). *Survey of face detection and recognition methods* [online]. Dostupné z: [https://www.researchgate.net/publication/237049645\\_Survey\\_of\\_face\\_detection\\_and\\_recognition\\_methods](https://www.researchgate.net/publication/237049645_Survey_of_face_detection_and_recognition_methods)
- [22] HOLČÍK, Jiří. *Analýza a klasifikace dat* [online]. Brno: AKADEMICKÉ NAKLADATELSTVÍ CERM, 2012 [cit. 2020-01-04]. ISBN 978-80-7204-793-2. Dostupné z: <https://www.iba.muni.cz/res/file/ucebnice/holcik-analyza-klasifikace-dat.pdf>
- [23] SONI, Devin. Supervised vs. Unsupervised Learning [online]. 2018, 22. Březen [cit. 2020-01-04]. Dostupné z: <https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d>
- [24] LACKO, Luboslav. Strojové učení: S učitelem i bez něj. *CIO Bussiness World* [online]. 2019 [cit. 2020-01-04]. Dostupné z: <https://businessworld.cz/bi-a-data/strojove-uceni-s-ucitelem-i-bez-nej-14868>
- [25] BRUNELLI, R. a T. POGGIO. *Face recognition: features versus templates* [online]. Oct 1993, , 1042 - 1052 [cit. 2020-01-05]. DOI: 10.1109/34.254061. ISSN 1939-3539. Dostupné z: <https://ieeexplore.ieee.org/document/254061/authors#authors>
- [26] DEEPAK GHIMIRE a JOONWHOAN LEE. *Geometric Feature-Based Facial Expression Recognition in Image Sequences Using Multi-Class AdaBoost and Support Vector Machines* [online]. 2013, , 21 [cit. 2020-01-05]. DOI: 10.3390/s130607714. Dostupné z: <https://www.mdpi.com/1424-8220/13/6/7714/html>
- [27] PARMAR, Divyarajsinh N. a Brijesh B. MEHTA. *Face Recognition Methods & Applications* [online]. Tech Wadhwani city, India, 2014, , 4 [cit. 2020-01-05]. ISSN 2229-6093. Dostupné z: [https://www.researchgate.net/publication/260483303\\_Face\\_Recognition\\_Methods\\_Applications](https://www.researchgate.net/publication/260483303_Face_Recognition_Methods_Applications)
- [28] VIJAYA LATA, Y. a Chandra Kiran Bharadwaj TUNGATHURTHI. *Facial Recognition using Eigenfaces by PCA* [online]. International Journal of Recent Trends in Engineering, 2009, , 4 [cit. 2020-01-05]. Dostupné z: [https://www.researchgate.net/profile/Dr\\_Govardhan/publication/228748710\\_Facial\\_recognition\\_using\\_eigenfaces\\_by\\_PCA/links/00b4952948d3d00b19000000.pdf](https://www.researchgate.net/profile/Dr_Govardhan/publication/228748710_Facial_recognition_using_eigenfaces_by_PCA/links/00b4952948d3d00b19000000.pdf)
- [29] GÜNEŞ, Funda. Principal Component Analysis for Dimensionality Reduction. *The SAS Data Science Blog* [online]. October 2015 [cit. 2020-01-05]. Dostupné z: <https://blogs.sas.com/content/subconsciousmusings/2015/10/26/principal-component-analysis-for-dimensionality-reduction/>

- [30] Classification of NIR spectra by Linear Discriminant Analysis in Python. *NIRPY Research: Linear Discriminant Analysis* [online]. 2020 [cit. 2020-01-05]. Dostupné z: <https://nirpyresearch.com/classification-nir-spectra-linear-discriminant-analysis-python/>
- [31] GÁNDARA, Agustín. Face Recognition with OpenCV. *MEDIUM* [online]. May 2018 [cit. 2020-01-05]. Dostupné z: <https://medium.com/@agustindev/face-recognition-with-opencv-280ec1213ffd>
- [32] CAO, Qi a Ruishan LIU. *Real-Time Face Tracking and Replacement* [online]. , 1-6 [cit. 2020-02-13]. Dostupné z: <https://web.stanford.edu/class/cs231m/projects/final-report-cao-liu.pdf>
- [33] DAN CASAS, Guix. *REAL-TIME FACE TRACKING METHODS* [online]. Barcelona, 2009 [cit. 2020-02-13]. Dostupné z: <https://dancasas.github.io/docs/pfc.pdf>. Universitat Autònoma de Barcelona.
- [34] HSL and HSV. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2020 [cit. 2020-02-15]. Dostupné z: [https://en.wikipedia.org/wiki/HSL\\_and\\_HSV](https://en.wikipedia.org/wiki/HSL_and_HSV)
- [35] LECUN, Yann, Koray KAVUKCUOGLU a Clement FARABET. Convolutional networks and applications in vision. *IEEE* [online]. Paris, France, 2010, 03 August [cit. 2020-03-15]. DOI: 10.1109/ISCAS.2010.5537907. Dostupné z: <https://ieeexplore.ieee.org/abstract/document/5537907>
- [36] JAROSLAV, Jirkovský. Metody Deep Learning k segmentaci obrazu [online]. 2017, Listopad, , 3 [cit. 2020-03-28]. Dostupné z: [http://automa.cz/Aton/FileRepository/pdf\\_articles/11149.pdf](http://automa.cz/Aton/FileRepository/pdf_articles/11149.pdf)
- [37] Yashwanth M (2020). Face-Recognition-by-CNN, GitHub [cit. 2020-02-15]. Dostupné z: <https://www.github.com/Yash0330/Face-Recognition-by-CNN>
- [38] AlexNet. The MathWorks, Inc. [online]. United States, 2020 [cit. 2020-03-28]. Dostupné z: <https://www.mathworks.com/help/deeplearning/ref/alexnet.html>
- [39] Medium: A Practical guide to ReLU [online]. 2017, 30. listopad [cit. 2020-03-28]. Dostupné z: <https://medium.com/@danqing/a-practical-guide-to-relu-b83ca804f1f7>
- [40] Medium: Understanding the basics of CNN with image classification [online]. 2019, 4. říjen [cit. 2020-03-28]. Dostupné z: <https://becominghuman.ai/understanding-the-basics-of-cnn-with-image-classification-7f3a9ddea8f9>
- [41] DETTMERS, Tim. Deep Learning in a Nutshell: History and Training: RMSprop. NVIDIA developer [online]. 2015, 16. prosinec [cit. 2020-03-30]. Dostupné z: <https://devblogs.nvidia.com/deep-learning-nutshell-history-training/>
- [42] SHARMA, Sagar. Epoch vs Batch Size vs Iterations. *MEDIUM* [online]. 2017, 23. srpen [cit. 2020-03-30]. Dostupné z: <https://towardsdatascience.com/epoch-vs-iterations-vs-batch-size-4dfb9c7ce9c9>
- [43] DAS, Siddharth. CNN Architectures: LeNet, AlexNet, VGG, GoogLeNet, ResNet and more.... *MEDIUM* [online]. 2017, 16. listopad [cit. 2020-03-30]. Dostupné z: <https://medium.com/analytics-vidhya/cnns-architectures-lenet-alexnet-vgg-googlenet-resnet-and-more-666091488df5>
- [44] MathWorks. Introducing Deep Learning with MATLAB [online]. , 14 [cit. 2020-03-28]. Dostupné z: [https://www.systematics.co.il/wp-content/uploads/Deep\\_Learning\\_ebook.pdf](https://www.systematics.co.il/wp-content/uploads/Deep_Learning_ebook.pdf)

## Seznam příloh

Tabulky výsledků jednotlivých testování detekce a klasifikace obličejů. Výsledky sítě natrénované základní trénovací databází jsou označeny modře a výsledky sítě natrénované rozšířenou databází jsou označeny oranžově.

- **Příloha I:** Výsledky testování vůči šumu: Gaussian a Spleckle
- **Příloha II:** Výsledky testování vůči šumu Salt & Pepper a vůči změně expozice
- **Příloha III:** Výsledky testování vůči změně rozlišení a směru osvětlení

# Příloha I

Zdroj dat	test_dataset_1		test_dataset_1		test_dataset_1		test_dataset_1		test_dataset_1		test_dataset_1		test_dataset_1		test_dataset_1		test_dataset_1	
	žádný	Gausův	žádný	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	
Typ šumu	0	0	0	0	0,00025	0,0005	0,00025	0,001	0,0005	0,001	0,0005	0,001	0,0005	0,001	0,0005	0,01	0,2	
Intenzita šumu	102	101	102	101	102	101	102	101	99	100	98	100	98	100	85	60	96	
Počet DO == 1	0	1	0	1	0	1	0	1	3	2	4	2	4	17	17	60	96	
Počet DO == 0	0	1	0	1	0	1	0	1	3	2	4	2	4	17	17	60	96	
Přesnost detekce %	100	99,01960784	100	99,01960784	100	99,01960784	100	97,05882353	98,03921569	96,07843137	83,33333333	41,17647059	5,8823252941	83,33333333	41,17647059	5,8823252941	5,8823252941	
Počet RO == 1	102	101	101	101	94	101	94	68	56	54	12	6	0	12	6	6	6	
Počet RO == 0	0	0	1	0	8	0	8	31	44	44	73	36	0	73	36	6	6	
Přesnost rozpoznání %	100	99,01960784	96,85413	94,33994	92,15686275	90,69649	76,43413	68,68686869	66,27436	55,10204082	14,11764706	14,28571429	0	43,64383	40,39050	0	0	
Skóre pro RO == 1 (%)	98,64082	97,36701	96,85413	94,33994	90,69649	76,43413	68,68686869	66,27436	55,10204082	14,11764706	14,28571429	40,39050	0	43,64383	40,39050	0	0	
Zdroj dat	žádný	Gausův	žádný	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	Gausův	
Typ šumu	0	0	0	0	0,00025	0,0005	0,00025	0,001	0,0005	0,001	0,0005	0,001	0,0005	0,001	0,0005	0,1	0,2	
Intenzita šumu	102	101	102	101	102	101	102	101	100	101	100	100	100	100	86	46	4	
Počet DO == 1	0	1	0	1	0	1	1	2	2	2	16	56	98	16	56	98	98	
Počet DO == 0	0	1	0	1	0	1	1	2	2	2	16	56	98	16	56	98	98	
Přesnost detekce %	100	99,01960784	100	99,01960784	99,01960784	99,01960784	98,03921569	98,03921569	98,03921569	98,03921569	84,31372549	45,08039322	3,921586827	84,31372549	45,08039322	3,921586827	3,921586827	
Počet RO == 1	102	101	102	100	100	100	96	82	60	51	19	9	0	19	9	0	0	
Počet RO == 0	0	0	0	1	5	1	5	18	40	49	67	37	4	67	37	4	4	
Přesnost rozpoznání %	100	100	100	99,00900909	99,00900909	99,00900909	95,04950495	82	60	51	22,09302326	19,56521739	0	22,09302326	19,56521739	0	0	
Skóre pro RO == 1 (%)	99,79924	98,59568	98,66432	98,07590	89,44597	89,44597	89,44597	74,98370	68,52942	68,88792	45,0862	28,35433	0	45,0862	28,35433	0	0	

Zdroj dat	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1
Typ šumu	Speckle	Speckle	Speckle	Speckle	Speckle	Speckle	Speckle	Speckle	Speckle	Speckle	Speckle
Intenzita šumu - var	0	0,00025	0,0005	0,0075	0,0025	0,001	0,0075	0,01	0,05	0,1	0,2
Počet DO = 1	102	101	101	101	101	101	101	101	100	100	93
Počet DO = 0	0	1	1	1	1	1	1	1	2	2	9
Přesnost detekce %	100	99,01960784	99,01960784	99,01960784	99,01960784	99,01960784	99,01960784	98,03921569	98,03921569	98,03921569	91,17647059
Počet RO = 1	102	101	101	101	101	99	99	55	32	22	22
Počet RO = 0	0	0	0	0	0	0	0	2	45	68	71
Řesnost rozpoznání %	100	99,01960784	99,01960784	99,01960784	99,01960784	98,01980198	98,01980198	98,01980198	55	32	23,65591398
Skóre pro RO = 1 (%)	98,77581	98,70926	98,63815	98,58752	97,31479	99,23240	88,11010	64,17691	57,81059	45,9422	
Zdroj dat	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1
Typ šumu	Speckle	Speckle	Speckle	Speckle	Speckle	Speckle	Speckle	Speckle	Speckle	Speckle	Speckle
Intenzita šumu	0	var = 0,00025	var = 0,0005	var = 0,00075	var = 0,001	var = 0,0025	var = 0,005	var = 0,0075	var = 0,01	var = 0,1	var = 0,2
Počet DO = 1	102	101	102	102	102	102	102	101	102	99	94
Počet DO = 0	0	1	0	0	0	0	0	1	0	3	8
Přesnost detekce %	100	99,01960784	100	100	100	100	99,01960784	99,01960784	98,03921569	97,05882353	92,15686275
Počet RO = 1	102	101	102	102	102	101	99	99	20	36	20
Počet RO = 0	0	0	0	0	0	1	2	51	49	63	74
Řesnost rozpoznání %	100	100	100	100	100	99,01960784	98,01980198	96,07841317	49	36	21,27659574
Skóre pro RO = 1 (%)	99,79924	99,72644	99,66882	99,42886	99,54323	97,71966	96,36566	79,47402	76,86775	66,31050	

Příloha II

Zdroj dat Typ šumu Intenzita šumu Počet DO == 1 Počet DO == 0 Přesnost detekce % Počet RO == 1 Počet RO == 0 Přesnost rozpoznání % Skóre pro RO == 1 (%)	test_dataset_1 žádný	test_dataset_1 Salt and Pepper 0,00025	test_dataset_1 Salt and Pepper 0,0005	test_dataset_1 Salt and Pepper 0,00075	test_dataset_1 Salt and Pepper 0,001	test_dataset_1 Salt and Pepper 0,0025	test_dataset_1 Salt and Pepper 0,005	test_dataset_1 Salt and Pepper 0,0075	test_dataset_1 Salt and Pepper 0,01	test_dataset_1 Salt and Pepper 0,05	test_dataset_1 Salt and Pepper 0,1	test_dataset_1 Salt and Pepper 0,2
	102	102	102	102	102	102	102	101	102	100	94	56
	0	0	0	0	0	0	0	1	0	2	8	46
	100	100	100	100	100	100	100	99,01960784	100	98,03921569	92,15686275	54,90196078
	102	102	102	102	102	93	78	75	73	30	18	5
	0	0	0	0	0	9	24	26	29	70	76	51
	100	100	100	100	100	91,17647059	76,47058824	74,25742574	71,56862745	30	19,14893617	8,928571429
	98,64082	96,99344	95,59586	92,98041	90,64543	82,79958	79,47894	73,33651	70,67222	44,47830	45,39622	38,16820
	test_dataset_1 žádný	test_dataset_1 Salt and Pepper 0,00025	test_dataset_1 Salt and Pepper 0,0005	test_dataset_1 Salt and Pepper 0,00075	test_dataset_1 Salt and Pepper 0,001	test_dataset_1 Salt and Pepper 0,0025	test_dataset_1 Salt and Pepper 0,005	test_dataset_1 Salt and Pepper 0,0075	test_dataset_1 Salt and Pepper 0,01	test_dataset_1 Salt and Pepper 0,05	test_dataset_1 Salt and Pepper 0,1	test_dataset_1 Salt and Pepper 0,2
	102	102	102	102	102	102	101	102	101	99	95	55
Zdroj dat Typ šumu Intenzita šumu Počet DO == 1 Počet DO == 0 Přesnost detekce % Počet RO == 1 Počet RO == 0 Přesnost rozpoznání % Skóre pro RO == 1 (%)	test_dataset_1 žádný	test_dataset_1 Salt and Pepper 0,00025	test_dataset_1 Salt and Pepper 0,0005	test_dataset_1 Salt and Pepper 0,00075	test_dataset_1 Salt and Pepper 0,001	test_dataset_1 Salt and Pepper 0,0025	test_dataset_1 Salt and Pepper 0,005	test_dataset_1 Salt and Pepper 0,0075	test_dataset_1 Salt and Pepper 0,01	test_dataset_1 Salt and Pepper 0,05	test_dataset_1 Salt and Pepper 0,1	test_dataset_1 Salt and Pepper 0,2
	102	102	102	102	102	102	101	102	101	99	95	55
	0	0	0	0	0	0	1	0	1	3	7	47
	100	100	100	100	100	100	99,01960784	100	99,01960784	97,05882353	93,1372549	53,92156863
	102	102	102	102	102	96	85	70	68	34	22	6
	0	0	0	0	0	6	16	32	33	65	73	49
	100	100	100	100	100	94,11764706	84,15841584	68,62745098	67,32673267	34,34343434	23,15789474	10,90909091
	99,79924	99,55011	99,22200	98,63880	97,10890	91,97665	86,26301	85,21334	79,09637	63,53897	54,57132	42,99350
	test_dataset_1 žádný	test_dataset_1 Salt and Pepper 0,00025	test_dataset_1 Salt and Pepper 0,0005	test_dataset_1 Salt and Pepper 0,00075	test_dataset_1 Salt and Pepper 0,001	test_dataset_1 Salt and Pepper 0,0025	test_dataset_1 Salt and Pepper 0,005	test_dataset_1 Salt and Pepper 0,0075	test_dataset_1 Salt and Pepper 0,01	test_dataset_1 Salt and Pepper 0,05	test_dataset_1 Salt and Pepper 0,1	test_dataset_1 Salt and Pepper 0,2
	102	102	102	102	102	102	101	102	101	99	95	55
Zdroj dat Změna expozice Počet DO == 1 Počet DO == 0 Přesnost detekce % Počet RO == 1 Počet RO == 0 Přesnost rozpoznání % Skóre pro RO == 1 (%)	test_dataset_1 -125	test_dataset_1 -100	test_dataset_1 -75	test_dataset_1 -50	test_dataset_1 -25	test_dataset_1 0	test_dataset_1 +25	test_dataset_1 +50	test_dataset_1 +75	test_dataset_1 +100	test_dataset_1 +125	test_dataset_1 +150
	41	79	100	101	101	102	102	102	102	102	102	102
	61	23	2	1	1	0	0	0	0	0	0	0
	40,19607843	77,45098039	98,03921569	99,01960784	99,01960784	100	100	100	100	100	100	100
	6	26	57	92	101	102	102	102	99	91	79	40
	35	53	43	9	0	0	0	0	3	11	23	62
	14,63414634	32,91139241	57	91,08910891	100	100	100	100	97,05882353	89,21568627	77,45098039	39,21568627
	57,78683	70,36388	73,68154	82,82884	96,33248	98,64082	98,21107	94,90655	90,11908	85,30009	71,08287	64,13733
	test_dataset_1 -125	test_dataset_1 -100	test_dataset_1 -75	test_dataset_1 -50	test_dataset_1 -25	test_dataset_1 0	test_dataset_1 +25	test_dataset_1 +50	test_dataset_1 +75	test_dataset_1 +100	test_dataset_1 +125	test_dataset_1 +150
	41	79	100	101	101	102	102	102	102	102	102	97
Zdroj dat Změna expozice Počet DO == 1 Počet DO == 0 Přesnost detekce % Počet RO == 1 Počet RO == 0 Přesnost rozpoznání % Skóre pro RO == 1 (%)	test_dataset_1 -125	test_dataset_1 -100	test_dataset_1 -75	test_dataset_1 -50	test_dataset_1 -25	test_dataset_1 0	test_dataset_1 +25	test_dataset_1 +50	test_dataset_1 +75	test_dataset_1 +100	test_dataset_1 +125	test_dataset_1 +150
	41	79	100	101	101	102	102	102	102	102	102	97
	61	23	2	1	1	0	0	0	0	0	0	5
	40,19607843	77,45098039	98,03921569	99,01960784	99,01960784	100	100	100	100	100	100	95,0803922
	8	31	74	94	101	102	102	101	98	87	77	38
	33	48	26	7	0	0	0	1	4	15	25	59
	19,51219512	39,24050633	74	93,06930693	100	100	100	99,01960784	96,07843137	85,2941765	75,49019608	39,17525773
	74,13213	74,92542	77,56088	91,93037	98,93255	99,78767	99,45643	97,62736	93,48852	88,20106	82,42651	70,35468
	test_dataset_1 -125	test_dataset_1 -100	test_dataset_1 -75	test_dataset_1 -50	test_dataset_1 -25	test_dataset_1 0	test_dataset_1 +25	test_dataset_1 +50	test_dataset_1 +75	test_dataset_1 +100	test_dataset_1 +125	test_dataset_1 +150
	41	79	100	101	101	102	102	102	102	102	102	97

Příloha III

Zdroj dat	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1
Zmenšení	1	0,875	0,75	0,625	0,5	0,375	0,25
Rozlišení	640x480	560x420	480x360	400x300	320x240	240x180	160x120
Počet DO == 1	102	102	102	102	101	89	27
Počet DO == 0	0	0	0	0	1	13	75
Přesnost detekce %	100	100	100	100	99,01960784	87,25490196	26,47058824
Počet RO == 1	102	102	102	102	101	85	26
Počet RO == 0	0	0	0	0	0	4	1
Přesnost rozpoznání %	100	100	100	100	100	95,50561798	96,2962963
Skóre pro RO == 1 [%]	98,64082	98,28864	98,25185	97,52397	96,79726	89,08489	85,66669

Zdroj dat	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1	test_dataset_1
Zmenšení	1	0,875	0,75	0,625	0,5	0,375	0,25
Rozlišení	640x480	560x420	480x360	400x300	320x240	240x180	160x120
Počet DO == 1	102	102	102	102	101	89	27
Počet DO == 0	0	0	0	0	1	13	75
Přesnost detekce %	100	100	100	100	99,01960784	87,25490196	26,47058824
Počet RO == 1	102	102	102	102	101	86	26
Počet RO == 0	0	0	0	0	0	3	1
Přesnost rozpoznání %	100	100	100	100	100	96,62921348	96,2962963
Skóre pro RO == 1 [%]	99,78767	99,76592	99,75050	99,48066	99,06673	95,90384	91,90908

Zdroj dat	test_dataset_1	test_dataset_2	test_dataset_3	test_dataset_4
Osvětlení	high	low	split	triangle
Počet DO == 1	102	102	102	102
Počet DO == 0	0	0	0	0
Přesnost detekce %	100	100	100	100
Počet RO == 1	102	54	44	53
Počet RO == 0	0	48	58	49
Přesnost rozpoznání %	100	52,94117647	43,1372549	51,96078431
Skóre pro RO == 1 [%]	98,64082	62,06245	62,20807	63,52483

Zdroj dat	test_dataset_1	test_dataset_2	test_dataset_3	test_dataset_4
Osvětlení	high	low	split	triangle
Počet DO == 1	102	102	102	102
Počet DO == 0	0	0	0	0
Přesnost detekce %	100	100	100	100
Počet RO == 1	102	102	85	102
Počet RO == 0	0	0	17	0
Přesnost rozpoznání %	100	100	83,33333333	100
Skóre pro RO == 1 [%]	99,79924	98,84597	98,53271	98,39077